

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:36:24 ; Search time 8.95784 Seconds  
(without alignments)  
476.010 Million cell updates/sec

Title: US-09-853-079-36

Perfect score: 102  
Sequence: 1 RCLSIKRXFXSXTTIXXXMMFFXXXXXFL 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

- 1: /SID82/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID82/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID82/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID82/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID82/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID82/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID82/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID82/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID82/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID82/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID82/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID82/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID82/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID82/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID82/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID82/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID82/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID82/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID82/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID82/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID82/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID82/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID82/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75	73.5	245	19 AAW56296	Babesia microti BM
2	75	73.5	245	19 AAW56297	Babesia microti BM
3	75	73.5	245	20 AAY24353	Babesia microti an
4	75	73.5	245	21 AAB30202	B. microti BMNI-17
5	75	73.5	245	23 ABB88948	Babesia microti an
6	72	70.6	32	20 AAY24357	Babesia microti an
7	72	70.6	32	21 AAB30206	B. microti BMNI-17
8	72	70.6	32	23 ABB88951	Babesia microti an
9	36	35.3	150	21 AAG34732	Arabidopsis thalia
10	36	35.3	195	21 AAG28927	Arabidopsis thalia

11	35	34.3	96	22	ABB10351	Human cDNA SEQ ID
12	35	34.3	96	22	AAU18132	Novel human uterin
13	35	34.3	96	22	AAU17045	Human novel secret
14	35	34.3	96	22	AAU19897	Novel human calciu
15	35	34.3	2560	22	ABB11404	Human FLAMINGO 1 h
16	35	34.3	2923	22	AAU07054	Human Flamingo pro
17	35	34.3	2923	23	AAU50866	Cadherin EGF LAG s
18	35	34.3	2936	23	AAU74826	Human REPTR 9 prot
19	35	34.3	2956	22	AAU07053	Human Flamingo pol
20	34.5	33.8	325	23	ABP29648	Streptococcus poly
21	34.5	33.8	325	23	ABP30060	Streptococcus poly
22	34	33.3	359	22	AAU89921	Human immune/haema
23	34	33.3	359	22	AAU16043	Human novel secret
24	34	33.3	441	22	AAU93742	Human polypeptide,
25	34	33.3	700	22	ABB69104	Drosophila melanog
26	33	32.4	208	22	ABG01527	Novel human diagno
27	33	32.4	393	23	ABB54120	Lactococcus lactis
28	33	32.4	481	22	ABB63417	Drosophila melanog
29	33	32.4	744	20	AAU37544	Amino acid sequenc
30	32	31.4	125	22	AAO04655	Human polypeptide
31	32	31.4	410	23	ABP29775	Streptococcus poly
32	32	31.4	410	23	ABP30545	Streptococcus poly
33	32	31.4	418	23	ABP27080	Streptococcus poly
34	32	31.4	512	23	ABB55010	Lactococcus lactis
35	32	31.4	982	22	ABG24086	Novel human diagno
36	32	31.4	2405	21	AAU42192	Human ORFX ORF1956
37	32	31.4	2707	18	AAU27161	Mouse receptor ME2
38	32	31.4	3014	22	AAU68533	Human novel cytoki
39	32	31.4	3014	22	AAU02196	Seven-pass transme
40	32	31.4	3028	22	AAE08586	Human NOV7 protein
41	32	31.4	3298	22	AAE03657	Human extracellular
42	31	30.4	32	22	ABB15816	Human nervous syst
43	31	30.4	58	23	ABP06010	Human ORFX protein
44	31	30.4	79	22	AAO06877	Human polypeptide
45	31	30.4	102	22	AAU91825	C glutamicum prote

ALIGNMENTS

RESULT 1  
AAW56296  
ID AAW56296 standard; Protein; 245 AA.

XX AAW56296;

XX 28-SEP-1998 (first entry)

XX Babesia microti BMNI-16 antigen sequence.

XX antigen: detection; diagnosis; vaccine; tick-borne disease;  
XX differentiation; Lyme disease; ehrlichiosis.

OS Babesia microti.

PN EP834567-A2.

PD 08-APR-1998.

XX 01-OCT-1997; 97EP-0117067.

PR 24-APR-1997; 97US-0845258.

PR 01-OCT-1996; 96US-0723142.

XX (CORI-) CORIXA CORP.

XX Houghton R, Lodes MJ, Reed SG, Sleath PR;

XX WPI; 1998-195465/18.

XX N-PSDB; AAV22746.

PT Polypeptides comprising Babesia microti antigens and their  
immunogenic fragments or epitopes - and related nucleic acid,

PT vectors, transformed cells and antibodies, useful for diagnosis of  
PT infection and in protective vaccines

XX  
PS Claim 1; Page 66; 113pp; English.

XX  
CC The sequence is that of a polypeptide comprising at least  
CC one antigenic portion of a Babesia microti antigen. It can be used  
CC to diagnose B. microti infection by detecting specific antibodies  
CC in usual immunoassays. Infection can also be diagnosed using:  
CC (a) primers or probes derived from the coding sequence, in  
CC standard amplification or hybridisation tests, or (b) using  
CC antibodies to detect the corresponding antigen. It is also  
CC useful in vaccines to protect against infection, especially  
CC when formulated with an adjuvant. The new diagnostic method  
CC allow rapid differentiation between B. microti infection and  
CC other tick-borne diseases (Lyme disease and ehrlichiosis) that  
CC have similar symptoms but require different treatments.

XX SQ Sequence 245 AA;

Query Match 73.5%; Score 75; DB 19; Length 245;  
Best Local Similarity 53.1%; Pred. No. 7.4e-08;  
Matches 17; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 RCLSIKRFXXSXTFTXIXXXMXFFXXXFL 32  
| | | | | | | | | | | | | | | | | |  
Db 84 RCLSIIRFYSSISTFTLIDFVMPFFTLFTVFL 115

RESULT 2

AAW56297 standard; Protein; 245 AA.

XX AC AAW56297;

XX DT 28-SEP-1998 (first entry)

XX DE Babesia microti BMNI-17 antigen sequence.

XX KW antigen; detection; diagnosis; vaccine; tick-borne disease;  
XX KW differentiation; Lyme disease; ehrlichiosis.

XX OS Babesia microti.

XX PN EP834567-A2.

XX PD 08-APR-1998.

XX PF 01-OCT-1997; 97EP-0117067.

XX PR 24-APR-1997; 97US-0845258.

XX PR 01-OCT-1996; 96US-0723142.

XX PA (CORI-) CORIXA CORP.

XX PI Houghton R, Lodes MJ, Reed SG, Sleath PR;

XX DR WPI; 1998-195465/18.

XX DR N-PSDB; AAV22747.

XX PT Polypeptides comprising Babesia microti antigens and their  
PT immunogenic fragments or epitopes - and related nucleic acid,  
PT vectors, transformed cells and antibodies, useful for diagnosis of  
PT infection and in protective vaccines

XX PS Claim 1; Page 67-68; 113pp; English.

XX CC The sequence is that of a polypeptide comprising at least

CC one antigenic portion of a Babesia microti antigen. It can be used  
CC to diagnose B. microti infection by detecting specific antibodies  
CC in usual immunoassays. Infection can also be diagnosed using:

CC (a) primers or probes derived from the coding sequence, in  
CC standard amplification or hybridisation tests, or (b) using

CC antibodies to detect the corresponding antigen. It is also  
CC useful in vaccines to protect against infection, especially  
CC when formulated with an adjuvant. The new diagnostic methods  
CC allow rapid differentiation between B. microti infection and  
CC other tick-borne diseases (Lyme disease and ehrlichiosis) that  
CC have similar symptoms but require different treatments.

XX SQ Sequence 245 AA;

Query Match 73.5%; Score 75; DB 19; Length 245;  
Best Local Similarity 53.1%; Pred. No. 7.4e-08;  
Matches 17; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 RCLSIKRFXXSXTFTXIXXXMXFFXXXFL 32  
| | | | | | | | | | | | | | | | | |  
Db 84 RCLSIIRFYSSISTFTLIDFVMPFFTLFTVFL 115

RESULT 3

AAW24353 standard; Protein; 245 AA.

XX AC AAW24353;

XX DT 16-SEP-1999 (first entry)

XX DE Babesia microti antigen BMNI-17.

XX KW Babesia microti; antigen; immunogen; diagnosis; infection; vaccine;  
XX KW immunity; detection.

XX OS Babesia microti.

XX PN WO9929869-A1.

XX PD 17-JUN-1999.

XX PF 11-DEC-1998; 98WO-US26437.

XX PR 11-DEC-1997; 97US-0990571.

XX PA (CORI-) CORIXA CORP.

XX PA (MAYO-) MAYO FOUNDATION.

XX PI Bruinsma E, Houghton R, Lodes MJ, Persing D, Reed SG;

XX PI Sleath PR;

XX DR WPI; 1999-385612/32.

XX DR N-PSDB; AAX90011.

XX PT New isolated Babesia microti polypeptides

XX PS Example 1; Page 84; 126pp; English.

XX CC The present invention describes isolated polypeptides comprising  
XX specific immunogenic portions of Babesia microti. AAX88983 to AAX88994  
XX encode specifically claimed B. microti immunogenic proteins, and  
XX AAY24327 to AAY24338 represent these proteins. B. microti polypeptides  
XX and nucleic acids can be used for detecting B. microti infections. They  
XX can also be used in vaccines for inducing protective immunity against B.  
XX microti infections. The present sequence represents a B. microti antigen.

XX SQ Sequence 245 AA;

Query Match 73.5%; Score 75; DB 20; Length 245;  
Best Local Similarity 53.1%; Pred. No. 7.4e-08;  
Matches 17; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 RCLSIKRFXXSXTFTXIXXXMXFFXXXFL 32  
| | | | | | | | | | | | | | | | | |  
Db 84 RCLSIIRFYSSISTFTLIDFVMPFFTLFTVFL 115

## RESULT 4

AAB30202  
ID AAB30202 standard; Protein: 245 AA.

XX AC AAB30202;

XX DT 12-FEB-2001 (first entry)

XX DE B. microti BMNI-17 antigen SEQ ID NO: 32.

XX KW Babesiosis; rodent parasite; tick-borne illness; antigen; disease diagnosis; disease prevention.

XX OS Babesia microti.

XX PN WC2000060090-A1.

XX PD 12-OCT-2000.

XX PF 05-APR-2000; 2000WO-US09136.

XX PR 05-APR-1999; 99US-0286488.

XX PR 17-MAR-2000; 2000US-0528784.

XX PA (CORI-) CORIXA CORP.

XX PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;

XX DR WPI: 2000-686939/67.

XX DR N-PSDB; AAC65094.

XX PT New polypeptides containing an antigenic portion of Babesia microti antigen and DNAs encoding the polypeptides, useful for diagnosing, treating or preventing B. microti infection, or for inducing protective immunity in a patient

XX PS Claim 13; Page 81-82; 118pp; English.

XX CC The present invention is related to the isolation of antigenic sequences from the rodent parasite Babesia microti. This organism is transmitted to humans by the same tick which transmits Lyme disease and ehrlichiosis. The organism causes a malaria-like infection known as babesiosis. The sequences identified by this invention can be used in the diagnosis, prevention and treatment of babesiosis.

XX SQ Sequence 245 AA;

Query Match 73.5%; Score 75; DB 21; Length 245;

Best Local Similarity 53.1%; Pred. No. 7.4e-08;

Matches 17; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 RCLSIKXRXSXTFTIXIXXXMXFFXXXXXFL 32

Db 84 RCLSIIRFYSSISTFTILIDFVMPFFTLFTYFL 115

## RESULT 5

ABB88948  
ID ABB88948 standard; Protein: 245 AA.

XX AC ABB88948;

XX DT 20-JUN-2002 (first entry)

XX DE Babesia microti antigen BMNI-17.

XX KW Protozoacide; vaccine; antigen; antigenic epitope; infection.

XX OS Babesia microti.

XX PN WC200185947-A2.

XX PD 15-NOV-2001.

XX

PF 09-MAY-2001; 2001WO-US15192.

XX

PR 10-MAY-2000; 2000US-0569098.

PR

27-JUN-2000; 2000US-0605724.

PR

07-SEP-2000; 2000US-0656888.

PR

10-OCT-2000; 2000US-0685436.

PR

13-DEC-2000; 2000US-0737178.

PR

26-FEB-2001; 2001US-0794764.

XX

PA (CORI-) CORIXA CORP.

XX

PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ; Secrist H;

XX

DR WPI: 2002-216691/27.

DR

N-PSDB; ABL89338.

XX

PS Claim 13; Page 87; 195pp; English.

XX

CC The present invention relates to novel Babesia microti antigens and their coding sequences. The B. microti antigens, antigenic epitopes of such antigens, and compositions comprising such antigens are useful for diagnosing and treating B. microti infection. The compositions are especially useful for enhancing immune response against B. microti infection. The present sequence was used to illustrate the invention.

XX

SQ Sequence 245 AA;

Query Match 73.5%; Score 75; DB 23; Length 245;

Best Local Similarity 53.1%; Pred. No. 7.4e-08;

Matches 17; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 RCLSIKXRXSXTFTIXIXXXMXFFXXXXXFL 32

Db 84 RCLSIIRFYSSISTFTILIDFVMPFFTLFTYFL 115

## RESULT 6

AAY24357

ID AAY24357 standard; peptide; 32 AA.

XX AC AAY24357;

XX

DT 16-SEP-1999 (first entry)

XX

DE Babesia microti antigen BMNI-17 degenerate repeat sequence.

XX

Babesia microti; antigen; immunogen; diagnosis; infection; vaccine; immunity; detection.

KW

Babesia microti.

XX

XX Key Location/Qualifiers

FT Misc-difference 6 /label= Met, Ile

FT

FT Misc-difference 9 /label= Tyr, Ser

FT

FT Misc-difference 10 /label= Ser, Phe

FT

FT Misc-difference 12 /label= Leu, Ile

FT

FT Misc-difference 13 /label= Pro, Ser, Leu

FT

FT Misc-difference 17 /label= Leu, Arg

FT

FT Misc-difference 19 /label= Glu, Asp, Gly

FT

FT Misc-difference 20

FT Misc-difference 21 /label= Ile, Phe  
 FT Misc-difference 23 /label= Ala, Val  
 FT Misc-difference 26 /label= Leu, Pro  
 FT Misc-difference 27 /label= Met, Thr  
 FT Misc-difference 28 /label= Ser, Leu  
 FT Misc-difference 29 /label= Val, Phe  
 FT Misc-difference 30 /label= Thr, Ile  
 FT Misc-difference 30 /label= Cys, Tyr  
 PN WO9929869-A1.  
 XX  
 PD 17-JUN-1999.  
 XX  
 PF 11-DEC-1998; 98WO-US26437.  
 XX  
 PR 11-DEC-1997; 97US-0990571.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PA (MAYO-) MAYO FOUNDATION.  
 XX  
 PI Bruinsma E, Houghton R, Lodes MJ, Persing D, Reed SG;  
 PI Sleath PR;  
 XX  
 DR WPI; 1999-385612/32.  
 XX  
 PT New isolated Babesia microti polypeptides  
 PS Example 1; Page 90; 126pp; English.  
 XX  
 CC The present invention describes isolated polypeptides comprising  
 CC specific immunogenic portions of Babesia microti. AAX88983 to AAX88994  
 CC encode specifically claimed B. microti immunogenic proteins, and  
 CC AAY24327 to AAY24338 represent these proteins. B. microti polypeptides  
 CC and nucleic acids can be used for detecting B. microti infections. They  
 CC can also be used in vaccines for inducing protective immunity against B.  
 CC microti infections. The present sequence represents a B. microti antigen  
 CC BMNI-17 degenerate repeat sequence.  
 XX  
 SQ Sequence 32 AA;  
 Query Match 70.6%; Score 72; DB 20; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-08;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RCLSIKRFXXSXXTFIXIXXXMXFFXXXXXFL 32  
 Db 1 RCLSIKRFXXSXXTFIXIXXXMXFFXXXXXFL 32  
 RESULT 7  
 AAB30206  
 ID AAB30206 standard; Peptide; 32 AA.  
 XX  
 AC AAB30206;  
 XX  
 DT 12-FEB-2001 (first entry)  
 XX  
 DE B. microti BMNI-17 antigen repeat sequence SEQ ID NO: 36.  
 XX  
 KW Babesiosis; rodent parasite; tick-borne illness; antigen;  
 KW disease diagnosis; disease prevention.  
 OS Babesia microti.  
 XX  
 PN WO200060090-A1.  
 XX

FT Misc-difference 21 /label= Ile, Phe  
 FT Misc-difference 23 /label= Ala, Val  
 FT Misc-difference 26 /label= Leu, Pro  
 FT Misc-difference 27 /label= Met, Thr  
 FT Misc-difference 28 /label= Ser, Leu  
 FT Misc-difference 29 /label= Val, Phe  
 FT Misc-difference 30 /label= Thr, Ile  
 FT Misc-difference 30 /label= Cys, Tyr  
 PN WO9929869-A1.  
 XX  
 PD 17-JUN-1999.  
 XX  
 PF 11-DEC-1998; 98WO-US26437.  
 XX  
 PR 11-DEC-1997; 97US-0990571.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PA (MAYO-) MAYO FOUNDATION.  
 XX  
 PI Bruinsma E, Houghton R, Lodes MJ, Persing D, Reed SG;  
 PI Sleath PR;  
 XX  
 DR WPI; 1999-385612/32.  
 XX  
 PT New isolated Babesia microti polypeptides  
 PS Example 1; Page 90; 126pp; English.  
 XX  
 CC The present invention describes isolated polypeptides comprising  
 CC specific immunogenic portions of Babesia microti. AAX88983 to AAX88994  
 CC encode specifically claimed B. microti immunogenic proteins, and  
 CC AAY24327 to AAY24338 represent these proteins. B. microti polypeptides  
 CC and nucleic acids can be used for detecting B. microti infections. They  
 CC can also be used in vaccines for inducing protective immunity against B.  
 CC microti infections. The present sequence represents a B. microti antigen  
 CC BMNI-17 degenerate repeat sequence.  
 XX  
 SQ Sequence 32 AA;  
 Query Match 70.6%; Score 72; DB 20; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-08;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RCLSIKRFXXSXXTFIXIXXXMXFFXXXXXFL 32  
 Db 1 RCLSIKRFXXSXXTFIXIXXXMXFFXXXXXFL 32  
 RESULT 7  
 AAB30206  
 ID AAB30206 standard; Peptide; 32 AA.  
 XX  
 AC AAB30206;  
 XX  
 DT 12-FEB-2001 (first entry)  
 XX  
 DE B. microti BMNI-17 antigen repeat sequence SEQ ID NO: 36.  
 XX  
 KW Babesiosis; rodent parasite; tick-borne illness; antigen;  
 KW disease diagnosis; disease prevention.  
 OS Babesia microti.  
 XX  
 PN WO200060090-A1.  
 XX

PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-US09136.  
 XX  
 PR 05-APR-1999; 99US-0286488.  
 PR 17-MAR-2000; 2000US-0528784.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Read SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;  
 XX  
 DR WPI; 2000-686939/67.  
 XX  
 PT New polypeptides containing an antigenic portion of Babesia microti  
 PT antigen and DNAs encoding the polypeptides, useful for diagnosing,  
 PT treating or preventing B. microti infection, or for inducing protective  
 PT immunity in a patient  
 XX  
 PS Claim 6; Page 85; 118pp; English.  
 XX  
 CC The present invention is related to the isolation of antigenic sequences  
 CC from the rodent parasite Babesia microti. This organism is transmitted to  
 CC humans by the same tick which transmits Lyme disease and ehrlichiosis.  
 CC The organism causes a malaria-like infection known as babesiosis. The  
 CC sequences identified by this invention can be used in the diagnosis,  
 CC prevention and treatment of babesiosis.  
 XX  
 SQ Sequence 32 AA;  
 Query Match 70.6%; Score 72; DB 21; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-08;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RCLSIKRFXXSXXTFIXIXXXMXFFXXXXXFL 32  
 Db 1 RCLSIKRFXXSXXTFIXIXXXMXFFXXXXXFL 32  
 RESULT 8  
 ABB88951  
 ID ABB88951 standard; Peptide; 32 AA.  
 XX  
 AC ABB88951;  
 XX  
 DT 20-JUN-2002 (first entry)  
 XX  
 DE Babesia microti antigen BMNI-17 epitope repeat.  
 XX  
 KW Protozoacide; vaccine; antigen; antigenic epitope; infection.  
 XX  
 OS Babesia microti.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 6 /label= Met, Ile  
 FT Misc-difference 9 /label= Tyr, Ser  
 FT Misc-difference 10 /label= Ser, Phe  
 FT Misc-difference 12 /label= Leu, Ile  
 FT Misc-difference 13 /label= Pro, Ser, Leu  
 FT Misc-difference 17 /label= Leu, Arg  
 FT Misc-difference 19 /label= Glu, Asp, Gly  
 FT Misc-difference 20 /label= Ile, Phe  
 FT Misc-difference 21 /label= Ala, Val  
 FT Misc-difference 23 /label= Leu, Pro



FT Misc-difference 26 /label= Met, Thr  
FT Misc-difference 27 /label= Ser, Leu  
FT Misc-difference 28 /label= val, Phe  
FT Misc-difference 29 /label= Thr, Ile  
FT Misc-difference 30 /label= Cys, Tyr  
XX  
PN WO200185947-A2.  
XX  
PN PD 15-NOV-2001.  
XX  
XX PF 09-MAY-2001; 2001WO-US15192.  
XX  
PR 10-MAY-2000; 2000US-0569098.  
PR 27-JUN-2000; 2000US-0605724.  
PR 07-SEP-2000; 2000US-0656688.  
PR 10-OCT-2000; 2000US-0685436.  
PR 13-DEC-2000; 2000US-0737178.  
PR 26-FEB-2001; 2001US-0794764.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;  
PI Secretist H;  
XX  
DR WPI; 2002-216691/27.  
XX  
PT New Babesia microti antigens, useful for diagnosing and treating B.  
PT microti infection, and as component of a composition for enhancing  
PT immune response against B. microti infections  
XX  
PS Claim 7; Page 91; 195pp; English.  
XX  
CC The present invention relates to novel Babesia microti antigens and their  
CC coding sequences. The B. microti antigens, antigenic epitopes of such  
CC antigens, and compositions comprising such antigens are useful for  
CC diagnosing and treating B. microti infection. The compositions are  
CC especially useful for enhancing immune response against B. microti  
CC infection. The present sequence was used to illustrate the invention.  
XX  
SQ Sequence 32 AA;  
Query Match 70.6%; Score 72; DB 23; Length 32;  
Best Local Similarity 100.0%; Pred. No. 4.1e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 RCLSIKXRFXXSXXTPIIXXXMXFFXXXFL 32  
DQ 1 RCLSIKXRFXXSXXTPIIXXXMXFFXXXFL 32  
RESULT 9  
AAG34732  
ID AAG34732 standard; Protein; 150 AA.  
XX  
AC AAG34732;  
XX  
XX 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 42309.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX

PD  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.

PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 08-OCT-1999; 99US-0158369.  
PR 12-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0161404.  
PR 23-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 35.3%; Score 36; DB 21; Length 150;

Best Local Similarity 43.8%; Pred. No. 7.7;

Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 LSIXRFXSXSTFIIXI 18

II: II I I I I

Db 5 LSVSRFMSSTATVISV 20

RESULT 10

AAG28927

ID AAG28927 standard; Protein; 195 AA.

XX

AC AAG28927;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 34328.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.

PR	07-OCT-1999;	99US-0158029.	PR	28-JUN-2000;	2000US-2148866P.
PR	08-OCT-1999;	99US-0158232.	PR	30-JUN-2000;	2000US-215135P.
PR	12-OCT-1999;	99US-0158369.	PR	07-JUL-2000;	2000US-216647P.
PR	13-OCT-1999;	99US-0159293.	PR	07-JUL-2000;	2000US-216880P.
PR	13-OCT-1999;	99US-0159294.	PR	11-JUL-2000;	2000US-217487P.
PR	13-OCT-1999;	99US-0159295.	PR	11-JUL-2000;	2000US-217496P.
PR	14-OCT-1999;	99US-0159329.	PR	14-JUL-2000;	2000US-218290P.
PR	14-OCT-1999;	99US-0159330.	PR	26-JUL-2000;	2000US-220963P.
PR	14-OCT-1999;	99US-0159331.	PR	26-JUL-2000;	2000US-220964P.
PR	14-OCT-1999;	99US-0159637.	PR	14-AUG-2000;	2000US-224518P.
PR	14-OCT-1999;	99US-0159638.	PR	14-AUG-2000;	2000US-224519P.
PR	18-OCT-1999;	99US-0159584.	PR	14-AUG-2000;	2000US-225213P.
PR	21-OCT-1999;	99US-0160741.	PR	14-AUG-2000;	2000US-225214P.
PR	21-OCT-1999;	99US-0160767.	PR	14-AUG-2000;	2000US-225266P.
PR	21-OCT-1999;	99US-0160768.	PR	14-AUG-2000;	2000US-225267P.
PR	21-OCT-1999;	99US-0160814.	PR	14-AUG-2000;	2000US-225268P.
PR	21-OCT-1999;	99US-0160815.	PR	14-AUG-2000;	2000US-225270P.
PR	22-OCT-1999;	99US-0160980.	PR	14-AUG-2000;	2000US-225447P.
PR	22-OCT-1999;	99US-0160981.	PR	14-AUG-2000;	2000US-225757P.
PR	22-OCT-1999;	99US-0160989.	PR	14-AUG-2000;	2000US-225758P.
PR	25-OCT-1999;	99US-0161404.	PR	14-AUG-2000;	2000US-225759P.
PR	25-OCT-1999;	99US-0161405.	PR	18-AUG-2000;	2000US-226279P.
PR	25-OCT-1999;	99US-0161406.	PR	22-AUG-2000;	2000US-226681P.
PR	26-OCT-1999;	99US-0161359.	PR	22-AUG-2000;	2000US-226688P.
PR	26-OCT-1999;	99US-0161360.	PR	22-AUG-2000;	2000US-227182P.
PR	26-OCT-1999;	99US-0161361.	PR	30-AUG-2000;	2000US-227009P.
PR	28-OCT-1999;	99US-0161920.	PR	30-AUG-2000;	2000US-228924P.
PR	28-OCT-1999;	99US-0161921.	PR	01-SEP-2000;	2000US-229287P.
PR	28-OCT-1999;	99US-0161992.	PR	01-SEP-2000;	2000US-229343P.
PR	28-OCT-1999;	99US-0161993.	PR	01-SEP-2000;	2000US-229344P.
PR	29-OCT-1999;	99US-0162142.	PR	01-SEP-2000;	2000US-229345P.
Query Match 35.3%; Score 36; DB 21; Length 195;					
Best Local Similarity 43.8%; Pred. No. 10;					
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps 0;					
Oy	3 LSIXRFXXSXXFXIXI 18				
Db	:          :				
5 LSVSRFMSSTATVISV 20					
RESULT 11					
ABBI0351	ID				
XX	ABBI0351 standard; Protein; 96 AA.				
AC	ABBI0351;				
XX	10-JAN-2002 (first entry)				
XX	Human cDNA SEQ ID NO: 659.				
DE	Human; gene therapy; neural disorder; immune system disorder;				
XX	muscular disorder; reproductive disorder; gastrointestinal disorder;				
KW	pulmonary disorder; cardiovascular disorder; renal disorder;				
KW	proliferative disorder; inflammation.				
KW	Homo sapiens.				
XX	WO200154474-A2.				
XX	02-AUG-2001.				
PD	17-JAN-2001; 2001WO-US01349.				
XX	31-JAN-2000; 2000US-179065P.				
PR	04-FEB-2000; 2000US-180628P.				
PR	24-FEB-2000; 2000US-184664P.				
PR	02-MAR-2000; 2000US-186350P.				
PR	16-MAR-2000; 2000US-189874P.				
PR	17-MAR-2000; 2000US-190076P.				
PR	18-APR-2000; 2000US-198123P.				
PR	19-MAY-2000; 2000US-205515P.				
PR	07-JUN-2000; 2000US-209467P.				

PR 20-OCT-2000; 2000US-2411786P.  
PR 20-OCT-2000; 2000US-2411787P.  
PR 20-OCT-2000; 2000US-2411808P.  
PR 20-OCT-2000; 2000US-2411809P.  
PR 20-OCT-2000; 2000US-2411826P.  
PR 01-NOV-2000; 2000US-244617P.  
PR 08-NOV-2000; 2000US-246417P.  
PR 08-NOV-2000; 2000US-246475P.  
PR 08-NOV-2000; 2000US-246476P.  
PR 08-NOV-2000; 2000US-246477P.  
PR 08-NOV-2000; 2000US-246478P.  
PR 08-NOV-2000; 2000US-246523P.  
PR 08-NOV-2000; 2000US-246524P.  
PR 08-NOV-2000; 2000US-246525P.  
PR 08-NOV-2000; 2000US-246526P.  
PR 08-NOV-2000; 2000US-246527P.  
PR 08-NOV-2000; 2000US-246528P.  
PR 08-NOV-2000; 2000US-246532P.  
PR 08-NOV-2000; 2000US-246609P.  
PR 08-NOV-2000; 2000US-246610P.  
PR 08-NOV-2000; 2000US-246611P.  
PR 08-NOV-2000; 2000US-246613P.  
PR 17-NOV-2000; 2000US-249207P.  
PR 17-NOV-2000; 2000US-249208P.  
PR 17-NOV-2000; 2000US-249209P.  
PR 17-NOV-2000; 2000US-249210P.  
PR 17-NOV-2000; 2000US-249211P.  
PR 17-NOV-2000; 2000US-249212P.  
PR 17-NOV-2000; 2000US-249213P.  
PR 17-NOV-2000; 2000US-249214P.  
PR 17-NOV-2000; 2000US-249215P.  
PR 17-NOV-2000; 2000US-249216P.  
PR 17-NOV-2000; 2000US-249217P.  
PR 17-NOV-2000; 2000US-249218P.  
PR 17-NOV-2000; 2000US-249244P.  
PR 17-NOV-2000; 2000US-249245P.  
PR 17-NOV-2000; 2000US-249264P.  
PR 17-NOV-2000; 2000US-249265P.  
PR 17-NOV-2000; 2000US-249297P.  
PR 17-NOV-2000; 2000US-249299P.  
PR 17-NOV-2000; 2000US-249300P.  
PR 01-DEC-2000; 2000US-250160P.  
PR 01-DEC-2000; 2000US-250391P.  
PR 05-DEC-2000; 2000US-251030P.  
PR 05-DEC-2000; 2000US-251988P.  
PR 05-DEC-2000; 2000US-256719P.  
PR 06-DEC-2000; 2000US-251479P.  
PR 08-DEC-2000; 2000US-251856P.  
PR 08-DEC-2000; 2000US-251868P.  
PR 08-DEC-2000; 2000US-251869P.  
PR 08-DEC-2000; 2000US-251989P.  
PR 11-DEC-2000; 2000US-254097P.  
PR 05-JAN-2001; 2001US-259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-476161/51.

N-PSDB; ABA06573.

Isolated nucleic acid molecule encoding an inflammation-associated polypeptide is used in preventing, treating or ameliorating a medical condition -

Claim 11; SEQ ID NO: 659; 859pp + Sequence Listing; English.

The present invention provides human cDNAs, proteins and related genomic DNAs. These can be used in the treatment of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders and inflammation. The present sequence is a protein of the invention.

XX SQ Sequence 96 AA;

Query Match 34.3%; Score 35; DB 22; Length 96;  
Best Local Similarity 54.5%; Pred. No. 8;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSLXREXXS 11  
II:I: II I  
Db 24 RCVSVLRFDSS 34

RESULT 12  
AAU18132

ID AAU18132 standard; Protein; 96 AA.

XX AC AAU18132;

XX DT 21-NOV-2001 (first entry)

XX DE Novel human uterine motility-association polypeptide #39.

XX DE Human; uterine motility-association disorder; uterus; pregnancy;  
XX KW labour; menstrual cycle; gene therapy.

XX OS Homo sapiens.

XX PN WO200155201-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01317.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 11-JUL-2000; 2000US-0218880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 18-AUG-2000; 2000US-0225759.

XX PR 22-AUG-2000; 2000US-0226861.

XX PR 22-AUG-2000; 2000US-0226868.

XX PR 23-AUG-2000; 2000US-0227182.

XX PR 30-AUG-2000; 2000US-0227009.

XX PR 01-SEP-2000; 2000US-0228924.

XX PR 01-SEP-2000; 2000US-0229287.

XX PR 01-SEP-2000; 2000US-0229343.

XX PR 01-SEP-2000; 2000US-0229344.

XX PR 01-SEP-2000; 2000US-0229345.

PR	05-SEP-2000;	2000US-022950-00;	2000US-022950-00;
PR	05-SEP-2000;	2000US-022951-03;	2000US-022951-03;
PR	06-SEP-2000;	2000US-023043-37;	2000US-023043-37;
PR	06-SEP-2000;	2000US-023043-38;	2000US-023043-38;
PR	08-SEP-2000;	2000US-023124-42;	2000US-023124-42;
PR	08-SEP-2000;	2000US-023124-43;	2000US-023124-43;
PR	08-SEP-2000;	2000US-023124-44;	2000US-023124-44;
PR	08-SEP-2000;	2000US-023141-13;	2000US-023141-13;
PR	08-SEP-2000;	2000US-023141-14;	2000US-023141-14;
PR	08-SEP-2000;	2000US-023208-00;	2000US-023208-00;
PR	08-SEP-2000;	2000US-023208-11;	2000US-023208-11;
PR	12-SEP-2000;	2000US-023196-68;	2000US-023196-68;
PR	14-SEP-2000;	2000US-023239-97;	2000US-023239-97;
PR	14-SEP-2000;	2000US-023239-98;	2000US-023239-98;
PR	14-SEP-2000;	2000US-023239-99;	2000US-023239-99;
PR	14-SEP-2000;	2000US-023240-01;	2000US-023240-01;
PR	14-SEP-2000;	2000US-023240-01;	2000US-023240-01;
PR	14-SEP-2000;	2000US-023306-31;	2000US-023306-31;
PR	14-SEP-2000;	2000US-023306-32;	2000US-023306-32;
PR	14-SEP-2000;	2000US-023306-35;	2000US-023306-35;
PR	21-SEP-2000;	2000US-023421-23;	2000US-023421-23;
PR	21-SEP-2000;	2000US-023424-74;	2000US-023424-74;
PR	25-SEP-2000;	2000US-023499-77;	2000US-023499-77;
PR	25-SEP-2000;	2000US-023499-78;	2000US-023499-78;
PR	26-SEP-2000;	2000US-023498-84;	2000US-023498-84;
PR	27-SEP-2000;	2000US-023568-34;	2000US-023568-34;
PR	27-SEP-2000;	2000US-023583-36;	2000US-023583-36;
PR	29-SEP-2000;	2000US-023632-27;	2000US-023632-27;
PR	29-SEP-2000;	2000US-023632-27;	2000US-023632-27;
PR	29-SEP-2000;	2000US-023636-77;	2000US-023636-77;
PR	29-SEP-2000;	2000US-023636-78;	2000US-023636-78;
PR	29-SEP-2000;	2000US-023636-79;	2000US-023636-79;
PR	29-SEP-2000;	2000US-023637-70;	2000US-023637-70;
PR	29-SEP-2000;	2000US-023637-70;	2000US-023637-70;
PR	29-SEP-2000;	2000US-023639-35;	2000US-023639-35;
PR	13-OCT-2000;	2000US-023993-35;	2000US-023993-35;
PR	13-OCT-2000;	2000US-023993-37;	2000US-023993-37;
PR	02-OCT-2000;	2000US-023680-02;	2000US-023680-02;
PR	02-OCT-2000;	2000US-023703-37;	2000US-023703-37;
PR	02-OCT-2000;	2000US-023703-38;	2000US-023703-38;
PR	02-OCT-2000;	2000US-023703-39;	2000US-023703-39;
PR	20-OCT-2000;	2000US-024096-60;	2000US-024096-60;
PR	20-OCT-2000;	2000US-024122-11;	2000US-024122-11;
PR	20-OCT-2000;	2000US-024178-55;	2000US-024178-55;
PR	20-OCT-2000;	2000US-024178-56;	2000US-024178-56;
PR	20-OCT-2000;	2000US-024178-57;	2000US-024178-57;
PR	20-OCT-2000;	2000US-024180-08;	2000US-024180-08;
PR	20-OCT-2000;	2000US-024180-09;	2000US-024180-09;
PR	20-OCT-2000;	2000US-024182-65;	2000US-024182-65;
PR	01-NOV-2000;	2000US-024461-77;	2000US-024461-77;
PR	08-NOV-2000;	2000US-024647-74;	2000US-024647-74;
PR	08-NOV-2000;	2000US-024647-75;	2000US-024647-75;
PR	08-NOV-2000;	2000US-024654-76;	2000US-024654-76;
PR	08-NOV-2000;	2000US-024654-77;	2000US-024654-77;
PR	08-NOV-2000;	2000US-024657-78;	2000US-024657-78;
PR	08-NOV-2000;	2000US-024658-23;	2000US-024658-23;
PR	08-NOV-2000;	2000US-024658-24;	2000US-024658-24;
PR	08-NOV-2000;	2000US-024659-25;	2000US-024659-25;
PR	08-NOV-2000;	2000US-024659-26;	2000US-024659-26;
PR	08-NOV-2000;	2000US-024659-27;	2000US-024659-27;
PR	08-NOV-2000;	2000US-024659-28;	2000US-024659-28;
PR	08-NOV-2000;	2000US-024660-09;	2000US-024660-09;
PR	08-NOV-2000;	2000US-024661-11;	2000US-024661-11;
PR	08-NOV-2000;	2000US-024661-13;	200

PR	17-NOV-2000; 2000US-0249216.	
PR	17-NOV-2000; 2000US-0249217.	
PR	17-NOV-2000; 2000US-0249218.	
PR	17-NOV-2000; 2000US-0249244.	
PR	17-NOV-2000; 2000US-0249245.	
PR	17-NOV-2000; 2000US-0249264.	
PR	17-NOV-2000; 2000US-0249265.	
PR	17-NOV-2000; 2000US-0249297.	
PR	17-NOV-2000; 2000US-0249299.	
PR	17-NOV-2000; 2000US-0249300.	
PR	01-DEC-2000; 2000US-0250160.	
PR	01-DEC-2000; 2000US-0250391.	
PR	05-DEC-2000; 2000US-0251030.	
PR	05-DEC-2000; 2000US-0251988.	
PR	05-DEC-2000; 2000US-0256719.	
PR	06-DEC-2000; 2000US-0251479.	
PR	08-DEC-2000; 2000US-0251856.	
PR	08-DEC-2000; 2000US-0251868.	
PR	08-DEC-2000; 2000US-0251869.	
PR	08-DEC-2000; 2000US-0251989.	
PR	08-DEC-2000; 2000US-0251990.	
PR	11-DEC-2000; 2000US-0254097.	
PR	05-JAN-2001; 2001US-0259678.	
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM;	
PI		
XX	WPI; 2001-488777/53.	
DR	N-PSDB; AAS28974.	
DR		
XX		
PT	Isolated polypeptide and nucleic acid molecules for treating,	
PT	preventing and/or prognosing disorders related to uterine motility	
PT	e.g. disorders associated with pregnancy and the menstrual cycle -	
XX		
XX	Claim 11; SEQ ID No 108; 524pp; English.	
XX		
CC	The present invention relates to the isolation of novel human	
CC	uterine motility-association polypeptides, and cDNA (AAS28936-AAS28994)	
CC	and genomic sequences encoding for these polypeptides.	
CC	The sequences of the invention are useful in the diagnosis,	
CC	treatment, prevention and/or prognosis of diseases associated	
CC	with uterine motility such as pregnancy and labour, and menstrual	
CC	disorders. The polynucleotide sequences of the invention are also	
CC	useful in gene therapy. AAU18094-AAU18152 represent novel human	
CC	uterine motility-association polypeptides.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences.	
XX		
XX	Sequence 96 AA;	
SQ		
	Query Match 34.3%; Score 35; DB 22; Length 96;	
	Best Local Similarity 54.5%; Pred. No. 8;	
	Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps	
QY	1 RCLSTXRFXXS 11	
	:	
Db	24 RCVSVLRDSS 34	
RESULT 13		
AAU17045		
ID	AAU17045 standard; Protein; 96 AA.	
XX		
XX	AAU17045;	
XX		
DT	07-NOV-2001 (first entry)	
XX		
XX	Human novel secreted protein, SEQ ID 286.	
DE		
XX	Human; immunosuppressive; antiarthritic; antirheumatic;	
KW	cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;	

KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnary; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.

XX Homo sapiens.

XX PN WO200155441-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01320.

XX PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.

PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.

PR 06-DEC-2000; 2000US-0251479.

PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Barash SC, Ruben SM;  
 PI  
 XX  
 DR WPI; 2001-476222/51.  
 DR N-PSDB; AAS26950.  
 XX  
 PT Novel polypeptides and polynucleotides useful as diagnostic reagents to  
 PT diagnose diseases or disorders associated with aberrant expression or  
 PT activity of polypeptides, for treating blood clotting disorder,  
 PT haemophilia  
 XX  
 PS Claim 11; SEQ ID No 286; 601pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. Antibodies to the proteins can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
 CC include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiodysplasia, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular disorders e.g. corneal infection, and many other  
 CC disorders listed in the specification. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present

Query Match 34.3%; Score 35; DB 22; Length 96;  
 Best Local Similarity 54.5%; Pred. No. 8;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSIKRPXKS 11  
 Db 24 RCVSVLRFDSS 34

RESULT 14  
 AAU19897  
 ID AAU19897 standard; Protein; 96 AA.  
 XX  
 AC AAU19897;  
 XX  
 DT 06-DEC-2001 (first entry)  
 XX  
 DE Novel human calcium-binding protein #6.  
 XX  
 KW Human; calcium-binding protein; calcium flux; neurological disease;  
 KW immune dysfunction; digestive disorder; neoplastic disease;  
 KW blood disorder; infectious disease; gene therapy; immunosuppressive;  
 KW antiarthritic; cytostatic; vasotropic; antibacterial; nootropic;  
 KW virucide.  
 OS Homo sapiens.  
 XX  
 XX WO200155304-A2.  
 PN

XX 02-AUG-2001.  
 PD 17-JAN-2001; 2001WO-US01302.  
 XX  
 PF 31-JAN-2000; 2000US-0179065.  
 XX 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 27-JUN-2000; 2000US-0209467.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0227009.  
 PR 01-SEP-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 21-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 25-SEP-2000; 2000US-0234977.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.





PA (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Drmanac RT;  
XX WPI; 2001-457740/49.  
DR N-PSDB; ABA08648.  
XX  
XX Human proteins and DNA encoding sequences useful for preventing,  
PT treating or ameliorating a medical condition in a mammalian subject  
PT e.g. arthritis and cancer -  
XX  
XX Claim 20: Page 179-181; 1963pp; English.  
XX  
XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
CC invention also relates to vectors and recombinant host cells comprising a  
CC nucleotide of the invention, methods of producing the novel polypeptides,  
CC antibodies against the polypeptides, methods of detecting the nucleotides  
CC or polypeptides in a sample, and methods of identifying compounds which  
CC bind to polypeptides of the invention. Although novel, many of the  
CC polypeptides of the invention have homology to known proteins, thereby  
CC giving an insight into their probable biological activities, and hence  
CC potential therapeutic applications. The polypeptides of the invention may  
CC have various activities, including cytokine, cell proliferation or cell  
CC differentiation activities; stem cell growth factor activity;  
CC haematopoiesis regulatory activity; tissue growth activity;  
CC immunomodulatory activity; activin- or inhibin-related activities;  
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
CC thrombolytic activities; receptor or ligand activities; or may be  
CC involved in oncogenesis, cancer cell proliferation or metastasis.  
CC Depending on their biological activities, polypeptides and nucleotides of  
CC the invention are useful for preventing, treating or ameliorating medical  
CC conditions, e.g., by protein or gene therapy. Such conditions include  
CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
CC vascular growth. Polypeptides involved with tissue regeneration and  
CC repair (or nucleic acids encoding them) may be used to promote wound  
CC healing (e.g., of burns, incisions and ulcers), while those with  
CC immunomodulatory activities may be used in the treatment of viral,  
CC bacterial and fungal infections in addition to immune disorders.  
CC Polypeptides with growth factor activity may be used in cell cultures to  
CC promote cell growth. For example, such polypeptides may be used to  
CC manipulate stem cells in culture to give rise to neuroepithelial cells  
CC that can be used to augment or replace cells damaged by illness,  
CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
CC may also be used in the diagnosis of the above conditions, and in drug  
CC screening techniques. The present sequence represents a novel human  
XX polypeptide of the invention.

SQ Sequence 2560 AA;

Query Match 34.3%; Score 35; DB 22; Length 2560;  
Best Local Similarity 54.5%; Pred. No. 2.2e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 1 RCLSIKRFXXS 11  
|||:|:|  
Db 875 RCVSLRFDSS 895

Search completed: July 16, 2003, 17:46:43  
Job time : 10.9578 secs





;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: SEED AND BERRY  
;; STREET: 6300 Columbia Center, 701 Fifth Avenue  
;; CITY: Seattle  
;; STATE: Washington  
;; COUNTRY: USA  
;; ZIP: 98104  
;;  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/845,258  
;; FILING DATE: 24-APR-1997  
;; CLASSIFICATION: A35  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Maki, David J.  
;; REGISTRATION NUMBER: 31,392  
;; REFERENCE/DOCKET NUMBER: 210121.426C1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (206) 622-4900  
;; TELEFAX: (206)682-6031  
;;  
;; INFORMATION FOR SEQ ID NO: 36:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 32 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 6  
;; OTHER INFORMATION: /note= "Residue can be either Met  
;; OTHER INFORMATION: or Ile"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 9  
;; OTHER INFORMATION: /note= "Residue can be either Tyr  
;; OTHER INFORMATION: or Ser"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 10  
;; OTHER INFORMATION: /note= "Residue can be either Ser  
;; OTHER INFORMATION: or Phe"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 12  
;; OTHER INFORMATION: /note= "Residue can be either Leu  
;; OTHER INFORMATION: or Ile"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 13  
;; OTHER INFORMATION: /note= "Residue can be Pro, Ser or  
;; OTHER INFORMATION: Leu"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 17  
;; OTHER INFORMATION: /note= "Residue can be either Leu  
;; OTHER INFORMATION: or Arg"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 19  
;; OTHER INFORMATION: /note= "Residue can be Glu, Asp or  
;; OTHER INFORMATION: Gly"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 20  
;; OTHER INFORMATION: /note= "Residue can be either Ile  
;; OTHER INFORMATION: or Phe"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 21  
;; OTHER INFORMATION: /note= "Residue can be either Ala

;; OTHER INFORMATION: or Val"  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 23  
;; OTHER INFORMATION: /note= "Residue can be either Leu  
;; OTHER INFORMATION: or Pro"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 26  
;; OTHER INFORMATION: /note= "Residue can be either Met  
;; OTHER INFORMATION: or Thr"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 27  
;; OTHER INFORMATION: /note= "Residue can be either Ser  
;; OTHER INFORMATION: or Leu"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 28  
;; OTHER INFORMATION: /note= "Residue can be either Val  
;; OTHER INFORMATION: or Phe"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 29  
;; OTHER INFORMATION: /note= "Residue can be either Thr  
;; OTHER INFORMATION: or Ile"  
;;  
;; FEATURE:  
;; NAME/KEY: Modified-site  
;; LOCATION: 30  
;; OTHER INFORMATION: /note= "Residue can be either Cys  
;; OTHER INFORMATION: or Tyr"  
;;  
;; US-08-845-258-36.  
;;  
Query Match 70.6%; Score 72; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 4.le-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
;;  
QY 1 RCLSIKRXFXSXXTFFIXIXXMXKFFXXXXXFL 32  
|||||  
DB 1 RCLSIKRXFXSXXTFFIXIXXMXKFFXXXXXFL 32  
|||||  
;;  
RESULT 6  
US-08-990-571-36  
; Sequence 36, Application US/08990571  
; Patent No. 6214971  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G. et al.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.  
; NUMBER OF SEQUENCES: 79  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/990,571  
; FILING DATE: 11-DEC-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.426C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206)682-6031

```

; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 6
; OTHER INFORMATION: /note= "Residue can be either Met
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: /note= "Residue can be either Tyr
; OTHER INFORMATION: or Ser"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 10
; OTHER INFORMATION: /note= "Residue can be either Ser
; OTHER INFORMATION: or Phe"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 12
; OTHER INFORMATION: /note= "Residue can be either Leu
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /note= "Residue can be Pro, Ser or
; OTHER INFORMATION: Leu"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 17
; OTHER INFORMATION: /note= "Residue can be either Leu
; OTHER INFORMATION: or Arg"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 19
; OTHER INFORMATION: /note= "Residue can be Glu, Asp or
; OTHER INFORMATION: Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 20
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: or Phe"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 21
; OTHER INFORMATION: /note= "Residue can be either Ala
; OTHER INFORMATION: or Val"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "Residue can be either Leu
; OTHER INFORMATION: or Pro"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "Residue can be either Met
; OTHER INFORMATION: or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
; OTHER INFORMATION: /note= "Residue can be either Ser
; OTHER INFORMATION: or Leu"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 28
; OTHER INFORMATION: /note= "Residue can be either Val
; OTHER INFORMATION: or Phe"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 30
; OTHER INFORMATION: /note= "Residue can be either Tyr
; OTHER INFORMATION: or Ser"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 32
; OTHER INFORMATION: /note= "Residue can be either Met
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 36
; OTHER INFORMATION: /note= "Residue can be either Thr
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 36
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: or Tyr"
; US-08-990-571-36
; Query Match 70.68; Score 72; DB 4; Length 32;
; Best Local Similarity 100.0%; Pred. No. 4.1e-08;
; Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; Qy 1 RCLSIKXRFXXSXXTFIXIXXXMXFXFXFXFXFL 32
; Db 1 RCLSIKXRFXXSXXTFIXIXXXMXFXFXFXFXFL 32
; RESULT 7
; US-08-723-142A-36
; Sequence 36, Application US/08723142A
; Patent No. 6306396
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF B. MICROTI INFECTION
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08723.142A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 6
; OTHER INFORMATION: /note= "Residue can be either Met
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: /note= "Residue can be either Tyr
; OTHER INFORMATION: or Ser"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 10
; OTHER INFORMATION: /note= "Residue can be either Met
; OTHER INFORMATION: or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 12
; OTHER INFORMATION: /note= "Residue can be either Leu
; OTHER INFORMATION: or Phe"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 13
; OTHER INFORMATION: /note= "Residue can be Pro, Ser or
; OTHER INFORMATION: Leu"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 17
; OTHER INFORMATION: /note= "Residue can be either Leu
; OTHER INFORMATION: or Arg"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 19
; OTHER INFORMATION: /note= "Residue can be Glu, Asp or
; OTHER INFORMATION: Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 20
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: or Phe"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 21
; OTHER INFORMATION: /note= "Residue can be either Ala
; OTHER INFORMATION: or Val"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "Residue can be either Leu
; OTHER INFORMATION: or Pro"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "Residue can be either Met
; OTHER INFORMATION: or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
; OTHER INFORMATION: /note= "Residue can be either Ser
; OTHER INFORMATION: or Leu"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 28
; OTHER INFORMATION: /note= "Residue can be either Val
; OTHER INFORMATION: or Phe"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 30
; OTHER INFORMATION: /note= "Residue can be either Tyr
; OTHER INFORMATION: or Ser"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 32
; OTHER INFORMATION: /note= "Residue can be either Met
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 36
; OTHER INFORMATION: /note= "Residue can be either Thr
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 36
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: or Tyr"
; US-08-990-571-36
; Query Match 70.68; Score 72; DB 4; Length 32;
; Best Local Similarity 100.0%; Pred. No. 4.1e-08;
; Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
; Qy 1 RCLSIKXRFXXSXXTFIXIXXXMXFXFXFXFXFL 32
; Db 1 RCLSIKXRFXXSXXTFIXIXXXMXFXFXFXFXFL 32
```

```
OTHER INFORMATION: /note= "Residue can be either Ser
OTHER INFORMATION: or Phe"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 12
OTHER INFORMATION: /note= "Residue can be either Leu
OTHER INFORMATION: or Ile"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 13
OTHER INFORMATION: /note= "Residue can be Pro, Ser or
OTHER INFORMATION: Leu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 17
OTHER INFORMATION: /note= "Residue can be either Leu
OTHER INFORMATION: or Arg"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 19
OTHER INFORMATION: /note= "Residue can be Glu, Asp or
OTHER INFORMATION: Gly"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 20
OTHER INFORMATION: /note= "Residue can be either Ile
OTHER INFORMATION: or Phe"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 21
OTHER INFORMATION: /note= "Residue can be either Ala
OTHER INFORMATION: or Val"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 23
OTHER INFORMATION: /note= "Residue can be either Leu
OTHER INFORMATION: or Pro"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 26
OTHER INFORMATION: /note= "Residue can be either Met
OTHER INFORMATION: or Thr"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 27
OTHER INFORMATION: /note= "Residue can be either Ser
OTHER INFORMATION: or Leu"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 28
OTHER INFORMATION: /note= "Residue can be either Val
OTHER INFORMATION: or Phe"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 29
OTHER INFORMATION: /note= "Residue can be either Thr
OTHER INFORMATION: or Ile"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 30
OTHER INFORMATION: /note= "Residue can be either Cys
OTHER INFORMATION: or Tyr"
US-08-723-142A-36
```

Query Match 70.6%; Score 72; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 4.1e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RCLSIKRFXXSXXTIFXIXXXMXFFXXXFL 32  
|||||  
Db 1 RCLSIKRFXXSXXTIFXIXXXMXFFXXXFL 32

```
RESULT 8
US-09-528-784A-36
; Sequence 36, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Babesia microti
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (6)...(6)
; OTHER INFORMATION: Xaa = Methionine or Isoleucine
; NAME/KEY: VARIANT
; LOCATION: (9)...(9)
; OTHER INFORMATION: Xaa = Tyrosine or Serine
; NAME/KEY: VARIANT
; LOCATION: (10)...(10)
; OTHER INFORMATION: Xaa = Serine or Phenylalanine
; NAME/KEY: VARIANT
; LOCATION: (12)...(12)
; OTHER INFORMATION: Xaa = Leucine or Isoleucine
; NAME/KEY: VARIANT
; LOCATION: (13)...(13)
; OTHER INFORMATION: Xaa = Proline, Serine or Leucine
; NAME/KEY: VARIANT
; LOCATION: (17)...(17)
; OTHER INFORMATION: Xaa = Leucine or Arginine
; NAME/KEY: VARIANT
; LOCATION: (19)...(19)
; OTHER INFORMATION: Xaa = Glutamic Acid, Aspartic Acid or Glycine
; NAME/KEY: VARIANT
; LOCATION: (20)...(20)
; OTHER INFORMATION: Xaa = Isoleucine or Phenylalanine
; NAME/KEY: VARIANT
; LOCATION: (21)...(21)
; OTHER INFORMATION: Xaa = Alanine or Valine
; NAME/KEY: VARIANT
; LOCATION: (23)...(23)
; OTHER INFORMATION: Xaa = Leucine or Proline
; NAME/KEY: VARIANT
; LOCATION: (26)...(26)
; OTHER INFORMATION: Xaa = Methionine or Threonine
; NAME/KEY: VARIANT
; LOCATION: (27)...(27)
; OTHER INFORMATION: Xaa = Serine or Leucine
; NAME/KEY: VARIANT
; LOCATION: (28)...(28)
; OTHER INFORMATION: Xaa = Valine or Phenylalanine
; NAME/KEY: VARIANT
; LOCATION: (29)...(29)
; OTHER INFORMATION: Xaa = Threonine or Isoleucine
; NAME/KEY: VARIANT
; LOCATION: (30)...(30)
; OTHER INFORMATION: Xaa = Cysteine ro Tyrosine
; OTHER INFORMATION: Xaa = Cysteine ro Tyrosine
US-09-528-784A-36
```

Query Match 70.6%; Score 72; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 4.1e-08;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404)873-8794  
; TELEFAX: (404)873-8795  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 335 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-182-616-4

Query Match 32.4%; Score 33; DB 4; Length 335;  
Best Local Similarity 50.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 RCLSIKRXFXSXXT 14  
DB 29 RCLQISSFANSWT 42

RESULT 12  
US-09-574-141A-64  
; Sequence 64, Application US/09574141A  
; Patent No. 6395490  
; GENERAL INFORMATION:  
; APPLICANT: Gonsalves, Dennis  
; APPLICANT: Meng, Baozhong  
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS  
; FILE REFERENCE: NUCLEIC ACIDS, PROTEINS, AND THEIR USES  
; FILE REFERENCE: 07678/035005  
; CURRENT APPLICATION NUMBER: US/09/574,141A  
; CURRENT FILING DATE: 2000-05-18  
; PRIOR APPLICATION NUMBER: 60/047,147  
; PRIOR FILING DATE: 1997-05-20  
; PRIOR APPLICATION NUMBER: 60/069,902  
; PRIOR FILING DATE: 1997-12-17  
; PRIOR APPLICATION NUMBER: 09/081,320  
; PRIOR FILING DATE: 1998-05-19  
; NUMBER OF SEQ ID NOS: 97  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 64  
; LENGTH: 379  
; TYPE: PRT  
; ORGANISM: Rupestris stem pitting associated virus  
US-09-574-141A-64

Query Match 32.4%; Score 33; DB 4; Length 379;  
Best Local Similarity 33.3%; Pred. No. 34;  
Matches 7; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 4 SIXRFXSXXTIFIXIXMXF 24  
DB 228 AIMRFSGEASTFLNTLANNLF 248

RESULT 13  
US-08-852-824-17  
; Sequence 17, Application US/08852824C  
; Patent No. 6060272  
; GENERAL INFORMATION:  
; APPLICANT: Li et al.  
; TITLE OF INVENTION: Human G-Protein Coupled Receptors  
; FILE REFERENCE: 1488.1220000  
; CURRENT APPLICATION NUMBER: US/08/852,824C  
; CURRENT FILING DATE: 1997-05-04  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: genomic

US-08-852-824-17

Query Match 30.4%; Score 31; DB 3; Length 348;  
Best Local Similarity 85.7%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CLSIXRF 8  
DB 121 CLSIDRF 127

RESULT 14  
US-08-383-750-4  
; Sequence 4, Application US/08383750  
; Patent No. 5744301  
; GENERAL INFORMATION:  
; APPLICANT: Birkenbach, Mark  
; APPLICANT: Kieff, Elliott  
; TITLE OF INVENTION: Epstein Barr Virus Induced Genes  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, N.W.,  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/383,750  
; FILING DATE: Herewith  
; CLASSIFICATION: 436  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fox, Samuel, L.  
; REGISTRATION NUMBER: 30,353  
; REFERENCE/DOCKET NUMBER: 0627.3300001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 361 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
US-08-383-750-4

Query Match 30.4%; Score 31; DB 1; Length 361;  
Best Local Similarity 85.7%; Pred. No. 83;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CLSIXRF 8  
DB 124 CLSIDRF 130

RESULT 15  
US-08-352-678-4  
; Sequence 4, Application US/08352678  
; Patent No. 6043351  
; GENERAL INFORMATION:  
; APPLICANT: Birkenbach, Mark  
; APPLICANT: Kieff, Elliott  
; TITLE OF INVENTION: EPSTEIN BARR VIRUS INDUCED GENES  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
; STREET: 600 Atlantic Avenue  
; CITY: Boston

STATE: MA  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/352,678  
FILING DATE: 30-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/980,518  
FILING DATE: 25-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Gates, Edward R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: B0801/7044  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 361 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-352-678-4

Query Match 30.4%; Score 31; DB 3; Length 361;  
Best Local Similarity 85.7%; Pred. No. 83;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CLSIXRF 8  
Db 124 CLSIDRF 130

Search completed: July 16, 2003, 17:50:59  
Job time : 3.914 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	75	73.5	245	9	US-09-286-488-32	Sequence 32, Appl
2	75	73.5	245	9	US-09-853-079-32	Sequence 32, Appl
3	75	73.5	245	10	US-09-737-178-32	Sequence 32, Appl
4	72	70.6	32	9	US-09-286-488-36	Sequence 36, Appl
5	72	70.6	32	9	US-09-853-079-36	Sequence 36, Appl
6	72	70.6	32	10	US-09-737-178-36	Sequence 36, Appl
7	35	34.3	96	9	US-10-073-865-108	Sequence 108, Appl
8	35	34.3	96	9	US-09-764-881-94	Sequence 94, Appl
9	35	34.3	96	10	US-09-764-853-659	Sequence 659, Appl
10	35	34.3	96	10	US-09-764-898-286	Sequence 286, Appl
11	35	34.3	2923	9	US-09-916-849-3	Sequence 3, Appl
12	35	34.3	2923	9	US-10-225-567A-524	Sequence 524, App
13	35	34.3	2923	10	US-09-788-711A-4	Sequence 4, Appl
14	35	34.3	2956	10	US-09-788-711A-2	Sequence 2, Appl
15	34	33.3	359	9	US-09-764-864-996	Sequence 996, Appl
16	33	32.4	1039	9	US-10-156-761-12232	Sequence 12232, A
17	32	31.4	1713	10	US-09-737-149-27	Sequence 27, Appl
18	32	31.4	3014	9	US-10-225-567A-444	Sequence 444, App
19	32	31.4	3014	10	US-09-737-149-2	Sequence 2, Appl

```
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C11
; CURRENT APPLICATION NUMBER: US/09/853,079
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-853-079-32

Query Match          73.5%; Score 75; DB 9; Length 245;
Best Local Similarity 53.1%; Pred. No. 2.1e-07;
Matches 17; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 RCLSIKRFXXXTFFIXXXMXFFXXXXXXFL 32
Db 84 RCLSIIRFYSSIFILIDFVMPFFTLFTYFL 115

RESULT 3
US-09-737-178-32
; Sequence 32, Application US/09737178
; Patent No. US20010029295A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-737-178-32

Query Match          73.5%; Score 75; DB 10; Length 245;
Best Local Similarity 53.1%; Pred. No. 2.1e-07;
Matches 17; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 RCLSIKRFXXXTFFIXXXMXFFXXXXXXFL 32
Db 84 RCLSIIRFYSSIFILIDFVMPFFTLFTYFL 115

RESULT 4
US-09-286-488-36
; Sequence 36, Application US/09286488
; Patent No. US20020169136A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C3
; CURRENT APPLICATION NUMBER: US/09/286,488
; CURRENT FILING DATE: 1999-04-05
```

```
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Babesia microti
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (6)...(6)
; OTHER INFORMATION: Xaa = Methionine or Isoleucine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (9)...(9)
; OTHER INFORMATION: Xaa = Tyrosine or Serine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (10)...(10)
; OTHER INFORMATION: Xaa = Serine or Phenylalanine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (12)...(12)
; OTHER INFORMATION: Xaa = Leucine or Isoleucine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (13)...(13)
; OTHER INFORMATION: Xaa = Proline, Serine or Leucine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (17)...(17)
; OTHER INFORMATION: Xaa = Leucine or Arginine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (19)...(19)
; OTHER INFORMATION: Xaa = Glutamic Acid, Aspartic Acid or Glycine
; NAME/KEY: VARIANT
; LOCATION: (20)...(20)
; OTHER INFORMATION: Xaa = Isoleucine or Phenylalanine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (21)...(21)
; OTHER INFORMATION: Xaa = Alanine or Valine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (23)...(23)
; OTHER INFORMATION: Xaa = Leucine or Proline
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (26)...(26)
; OTHER INFORMATION: Xaa = Methionine or Threonine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (27)...(27)
; OTHER INFORMATION: Xaa = Serine or Leucine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (28)...(28)
; OTHER INFORMATION: Xaa = Valine or Phenylalanine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (29)...(29)
; OTHER INFORMATION: Xaa = Threonine or Isoleucine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (30)...(30)
; OTHER INFORMATION: Xaa = Cysteine ro Tyrosine
US-09-286-488-36

Query Match          70.6%; Score 72; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RCLSIKRFXXXTFFIXXXMXFFXXXXXXFL 32
```

Db 1 RCLSIKRXSXXTIFIXIXXXMXFFXXXXXFL 32

## RESULT 5

US-09-853-079-36  
; Sequence 36, Application US/09853079  
; Publication No. US20030109689A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C11  
; CURRENT APPLICATION NUMBER: US/09/853.079  
; CURRENT FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 224  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 36  
; LENGTH: 32  
; TYPE: PRT  
; ORGANISM: Babesia microti  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (6)...(6)  
; OTHER INFORMATION: Xaa = Methionine or Isoleucine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (9)...(9)  
; OTHER INFORMATION: Xaa = Tyrosine or Serine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (10)...(10)  
; OTHER INFORMATION: Xaa = Serine or Phenylalanine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (12)...(12)  
; OTHER INFORMATION: Xaa = Leucine or Isoleucine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (13)...(13)  
; OTHER INFORMATION: Xaa = Proline, Serine or Leucine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (17)...(17)  
; OTHER INFORMATION: Xaa = Leucine or Arginine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (19)...(19)  
; OTHER INFORMATION: Xaa = Glutamic Acid, Aspartic Acid or Glycine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (20)...(20)  
; OTHER INFORMATION: Xaa = Isoleucine or Phenylalanine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (21)...(21)  
; OTHER INFORMATION: Xaa = Alanine or Valine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (23)...(23)  
; OTHER INFORMATION: Xaa = Leucine or Proline  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (26)...(26)  
; OTHER INFORMATION: Xaa = Methionine or Threonine  
; FEATURE:  
; NAME/KEY: VARIANT

; LOCATION: (27)...(27)  
; OTHER INFORMATION: Xaa = Serine or Leucine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (28)...(28)  
; OTHER INFORMATION: Xaa = Valine or Phenylalanine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (29)...(29)  
; OTHER INFORMATION: Xaa = Threonine or Isoleucine  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (30)...(30)  
; OTHER INFORMATION: Xaa = Cysteine ro Tyrosine  
US-09-853-079-36

Query Match 70.6%; Score 72; DB 9; Length 32;  
Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 RCLSIKRXSXXTIFIXIXXXMXFFXXXXXFL 32

Db 1 RCLSIKRXSXXTIFIXIXXXMXFFXXXXXFL 32

## RESULT 6

US-09-737-178-36  
; Sequence 36, Application US/09737178  
; Patent No. US20010029295A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C9  
; CURRENT APPLICATION NUMBER: US/09/737,178  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 36  
; LENGTH: 32  
; TYPE: PRT  
; ORGANISM: Babesia microti  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (6)...(6)  
; OTHER INFORMATION: Xaa = Methionine or Isoleucine  
; NAME/KEY: VARIANT  
; LOCATION: (9)...(9)  
; OTHER INFORMATION: Xaa = Tyrosine or Serine  
; NAME/KEY: VARIANT  
; LOCATION: (10)...(10)  
; OTHER INFORMATION: Xaa = Serine or Phenylalanine  
; NAME/KEY: VARIANT  
; LOCATION: (12)...(12)  
; OTHER INFORMATION: Xaa = Leucine or Isoleucine  
; NAME/KEY: VARIANT  
; LOCATION: (13)...(13)  
; OTHER INFORMATION: Xaa = Proline, Serine or Leucine  
; NAME/KEY: VARIANT  
; LOCATION: (17)...(17)  
; OTHER INFORMATION: Xaa = Leucine or Arginine  
; NAME/KEY: VARIANT  
; LOCATION: (19)...(19)  
; OTHER INFORMATION: Xaa = Glutamic Acid, Aspartic Acid or Glycine  
; NAME/KEY: VARIANT  
; LOCATION: (20)...(20)  
; OTHER INFORMATION: Xaa = Isoleucine or Phenylalanine

```
; NAME/KEY: VARIANT
; LOCATION: (21)...(21)
; OTHER INFORMATION: Xaa = Alanine or Valine
; NAME/KEY: VARIANT
; LOCATION: (23)...(23)
; OTHER INFORMATION: Xaa = Leucine or Proline
; NAME/KEY: VARIANT
; LOCATION: (26)...(26)
; OTHER INFORMATION: Xaa = Methionine or Threonine
; NAME/KEY: VARIANT
; LOCATION: (27)...(27)
; OTHER INFORMATION: Xaa = Serine or Leucine
; NAME/KEY: VARIANT
; LOCATION: (28)...(28)
; OTHER INFORMATION: Xaa = Valine or Phenylalanine
; NAME/KEY: VARIANT
; LOCATION: (29)...(29)
; OTHER INFORMATION: Xaa = Threonine or Isoleucine
; NAME/KEY: VARIANT
; LOCATION: (30)...(30)
; OTHER INFORMATION: Xaa = Cysteine ro Tyrosine
US-09-737-178-36

Query Match          70.68; Score 72; DB 10; Length 32;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RCLSIKRFXXSXXTFIXIXXXMXFFXXXXXFL 32
Db 1 RCLSIKRFXXSXXTFIXIXXXMXFFXXXXXFL 32

RESULT 7
US-10-073-865-108
; Sequence 108, Application US/10073865
; Publication No. US20030044904A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ09C1
; CURRENT APPLICATION NUMBER: US/10/073,865
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 154
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-073-865-108

Query Match          34.38; Score 35; DB 9; Length 96;
Best Local Similarity 54.5%; Pred. No. 8.2;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RCLSIKRFXXS 11
Db 24 RCVSVLRFDS 34

RESULT 8
US-09-764-881-94
; Sequence 94, Application US/09764881
; Publication No. US20030125246A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
```

```
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ07
; CURRENT APPLICATION NUMBER: US/09/764,881
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 94
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-881-94

Query Match          34.3%; Score 35; DB 9; Length 96;
Best Local Similarity 54.5%; Pred. No. 8.2;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RCLSIKRFXXS 11
Db 24 RCVSVLRFDS 34

RESULT 9
US-09-764-853-659
; Sequence 659, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ06
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 659
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (89)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (92)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-853-659

Query Match          34.3%; Score 35; DB 10; Length 96;
Best Local Similarity 54.5%; Pred. No. 8.2;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RCLSIKRFXXS 11
Db 24 RCVSVLRFDS 34

RESULT 10
US-09-764-898-286
; Sequence 286, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZ01
; CURRENT APPLICATION NUMBER: US/09/764,898
```

; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 311  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 286  
; LENGTH: 96  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (89)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (92)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-898-286

Query Match 34.3%; Score 35; DB 10; Length 96;  
Best Local Similarity 54.5%; Pred. No. 8.2;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSTXRFXXS 11  
||:|: || |  
Db 24 RCVSVLRFDS 34

RESULT 11  
US-09-916-849A-3  
; Sequence 3, Application US/09916849A  
; Publication No. US20030086934A1  
; GENERAL INFORMATION:  
; APPLICANT: Bostein, et al.  
; TITLE OF INVENTION: Basal Markers in Breast Cancer and Related Reagents  
; FILE REFERENCE: 2002850-0024  
; CURRENT APPLICATION NUMBER: US/09/916,849A  
; CURRENT FILING DATE: 2001-07-26  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 2923  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Cadherin EGF  
; OTHER INFORMATION: LAG Seven Pass G-Type Receptor 2  
US-09-916-849A-3

Query Match 34.3%; Score 35; DB 9; Length 2923;  
Best Local Similarity 54.5%; Pred. No. 2.1e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSTXRFXXS 11  
||:|: || |  
Db 1242 RCVSVLRFDS 1252

RESULT 12  
US-10-225-567A-524  
; Sequence 524, Application US/10225567A  
; Publication No. US20030113798A1  
; GENERAL INFORMATION:  
; APPLICANT: LifeSpan Biosciences  
; APPLICANT: Brown, Joseph P.  
; APPLICANT: Burner, Glenn A.  
; APPLICANT: Roush, Christine L.  
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS  
; FILE REFERENCE: 1920-4-4  
; CURRENT APPLICATION NUMBER: US/10/225,567A  
; CURRENT FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 60/257,144  
; PRIOR FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 2292

; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 524  
; LENGTH: 2923  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-225-567A-524

Query Match 34.3%; Score 35; DB 9; Length 2923;  
Best Local Similarity 54.5%; Pred. No. 2.1e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSTXRFXXS 11  
||:|: || |  
Db 1242 RCVSVLRFDS 1252

RESULT 13  
US-09-788-711A-4  
; Sequence 4, Application US/09788711A  
; Patent No. US20020058328A1  
; GENERAL INFORMATION:  
; APPLICANT: Tania Tamsin Testa  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30225  
; CURRENT APPLICATION NUMBER: US/09/788,711A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 0004196.2  
; PRIOR FILING DATE: 2000-02-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 2923  
; TYPE: PRT  
; ORGANISM: HOMO SAPIENS  
US-09-788-711A-4

Query Match 34.3%; Score 35; DB 10; Length 2923;  
Best Local Similarity 54.5%; Pred. No. 2.1e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSTXRFXXS 11  
||:|: || |  
Db 1242 RCVSVLRFDS 1252

RESULT 14  
US-09-788-711A-2  
; Sequence 2, Application US/09788711A  
; Patent No. US20020058328A1  
; GENERAL INFORMATION:  
; APPLICANT: Tania Tamsin Testa  
; TITLE OF INVENTION: NOVEL COMPOUNDS  
; FILE REFERENCE: GP-30225  
; CURRENT APPLICATION NUMBER: US/09/788,711A  
; CURRENT FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 0004196.2  
; PRIOR FILING DATE: 2000-02-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 2956  
; TYPE: PRT  
; ORGANISM: HOMO SAPIENS  
US-09-788-711A-2

Query Match 34.3%; Score 35; DB 10; Length 2956;  
Best Local Similarity 54.5%; Pred. No. 2.1e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSTXRFXXS 11  
||:|: || |  
Db 1242 RCVSVLRFDS 1252

RESULT 15  
US-09-764-864-996  
; Sequence 996, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764.864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 996  
; LENGTH: 359  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-864-996

Query Match 33.3%; Score 34; DB 10; Length 359;  
Best Local Similarity 37.5%; Pred. No. 45;  
Matches 6; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 1 RCLSIKRFXXSXTFI 16  
||: :| |||  
Db 105 RCIEGKFLKKHSTFI 120

Search completed: July 16, 2003, 18:08:35  
Job time : 6.2344 secs





A:Accession: PC4115  
A:Molecule type: DNA  
A:Residues: 1-312 <HAR>  
A:Cross-references: GB:U26453; GB:U26453; NID:g1216381  
C:Comment: This protein is a member of a gene complex involved in the biogenesis and assembly of flagellar protein or virulence factor.  
C:Genetics:  
A:Gene: flhB; TP0715  
C:Superfamily: flagellar biosynthetic protein flhB; flhB carboxyl-terminal homology  
C:Keywords: inner membrane

Query Match 35.3%; Score 36; DB 2; Length 376;  
Best Local Similarity 25.8%; Pred. No. 4.7;  
Matches 8; Conservative 4; Mismatches 19; Indels 0; Gaps 0;

QY 2 CLSIXRFXSXXTFIXIXXXMXFFXXXFL 32  
DB 75 CIGVLRFFFTTRATTASINTGWFVFRYFM 105

RESULT 3  
T20554  
hypothetical protein F07D3.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
A:Accession: T20554; T23359  
R:Berk, M.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19291  
A:Accession: T20554  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-208 <WIL>  
A:Cross-references: EMBL:Z71179; PIDN:CAA94889.1; GSPDB:GNO0023; CESP:F07D3.2  
A:Experimental source: clone F07D3  
R:Lloyd C.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19731  
A:Accession: T23359  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-208 <WIL>  
A:Cross-references: EMBL:Z70755; PIDN:CAA94786.1; GSPDB:GNO0023; CESP:F07D3.2  
A:Experimental source: clone K06A4  
C:Genetics:  
A:Gene: CESP:F07D3.2  
A:Map position: 5  
A:Introns: 37/3; 66/1; 127/2; 150/2; 182/2

Query Match 34.3%; Score 35; DB 2; Length 208;  
Best Local Similarity 40.0%; Pred. No. 4.6;  
Matches 6; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 CLSIXRFXSXXXTFI 16  
DB 17 CFSVQRFLLSYISFL 31

RESULT 4  
T20896  
hypothetical protein F14F11.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T20896  
R:Lightning, J.  
submitted to the EMBL Data Library, October 1995  
A:Reference number: Z19342  
A:Accession: T20896  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-776 <WIL>  
A:Cross-references: EMBL:Z54307; PIDN:CAA91092.1; GSPDB:GNO0020; CESP:F14F11.1  
A:Experimental source: clone F14F11

C:Genetics:  
A:Gene: CESP:F14F11.1  
A:Map position: 2  
A:Introns: 67/1; 93/3; 129/3; 169/3; 201/1; 248/3; 291/1; 325/3; 372/2; 440/3; 499/2;  
Query Match 34.3%; Score 35; DB 2; Length 776;  
Best Local Similarity 33.3%; Pred. No. 14;  
Matches 8; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

QY 2 CLSIXRFXSXXXTFIXIXXXMXFF 25  
DB 255 CPSPHKEVRSPPLTIIDVISTGAFF 278

RESULT 5  
H83754  
multidrug resistance protein BH0840 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C:Accession: H83754  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: H83754  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-106 <STO>  
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA804559.1; GSPDB:G  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH0840  
C:Superfamily: sugE protein

Query Match 33.3%; Score 34; DB 2; Length 106;  
Best Local Similarity 42.1%; Pred. No. 4.3;  
Matches 8; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 7 RFXSXXXTFIXIXXXMXFF 25  
DB 28 KFIPSVITFVIIASYVFF 46

RESULT 6  
HLHUR3  
T-cell surface glycoprotein CD1d precursor - human  
N:Alternate names: T-cell surface glycoprotein CD1.R3; thymocyte differentiation anti  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 21-Jul-2000  
C:Accession: S07715; A32217; I79469; A30599  
R:Calabi, F.; Jarvis, J.M.; Martin, L.; Milstein, C.  
Eur. J. Immunol. 19, 285-292, 1989  
A:Title: Two classes of CD1 genes  
A:Reference number: A30599; MUID:89196496; PMID:2467814  
A:Accession: S07715  
A:Molecule type: DNA  
A:Residues: 1-335 <CAL>  
A:Cross-references: EMBL:X14974; NID:g29767; PIDN:CAA33099.1; PID:g4539529  
A:Experimental source: clone lambda R3G1  
R:Balk, S.P.; Bleicher, P.A.; Terhorst, C.  
Proc. Natl. Acad. Sci. U.S.A. 86, 252-256, 1989  
A:Title: Isolation and characterization of a cDNA and gene coding for a fourth CD1 mo  
A:Reference number: A32217; MUID:89098892; PMID:2463622  
A:Accession: A32217  
A:Molecule type: mRNA  
A:Residues: 1-335 <BAL>  
A:Cross-references: GB:J04142; NID:g619799; PIDN:AAAS9673.1; PID:g619800  
R:Martin, L.H.; Calabi, F.; Milstein, C.  
Proc. Natl. Acad. Sci. U.S.A. 83, 9154-9158, 1986  
A:Title: Isolation of CD1 genes: a family of major histocompatibility complex-related  
A:Reference number: 159088; MUID:87067468; PMID:3097645  
A:Accession: I79469  
A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 203-295 <RES>  
 A:Cross-references: GB:M14664; NID:g180047; PIDN:AAA51935.1; PID:g180048  
 C:Genetics:  
 A:Gene: GDB:CD1D  
 A:Cross-references: GDB:l18764; OMIM:188410  
 A:Map position: lq22-lq23  
 A:Introns: 21/1; 110/1; 203/1; 296/1; 329/2  
 C:Superfamily: heterodimer with beta-2-microglobulin (see PIR:MGHUB2)  
 C:Keywords: glycoprotein; heterodimer; thymocyte; transmembrane protein  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-335/Product: T-cell surface glycoprotein CD1d #status predicted <MAT>  
 F:20-109/Domain: alpha-1 <EX1>  
 F:110-202/Domain: alpha-2 <EX2>  
 F:203-295/Domain: alpha-3 <EX3>  
 F:216-281/Domain: immunoglobulin homology <IMM>  
 F:302-321/Domain: transmembrane #status predicted <TM>  
 F:322-335/Domain: intracellular #status predicted <INT>  
 F:38,60,126,181/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:120-184,224-279/Disulfide bonds: #status predicted

Query Match 32.4%; Score 33; DB 1; Length 335;  
 Best Local Similarity 50.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 RCLSIKRFXXSXT 14  
 ||| | | | | | |  
 DB 29 RCLQISSFANSWT 42

RESULT 7  
 G86725  
 hypothetical protein yibE [imported] - Lactococcus lactis subsp. lactis (strain ILL1403)  
 C:Species: Lactococcus lactis subsp. lactis  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C:Accession: G86725  
 R:Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001  
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
 A:Reference number: A86625; MUID:21235186; PMID:11337471  
 A:Accession: G86725  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-393 <STO>  
 A:Cross-references: GB:AE005176; PID:g12723729; PIDN:AAK04905.1; GSPDB:GN00146  
 A:Experimental source: strain ILL1403  
 C:Genetics:  
 A:Gene: yibE

Query Match 32.4%; Score 33; DB 2; Length 393;  
 Best Local Similarity 26.1%; Pred. No. 22;  
 Matches 6; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 3 LSIXRFXXSXTFIXIXXXMXFF 25  
 : | | | | : : : |  
 DB 154 ISFERFRASIFLFLSLVLLLYF 176

RESULT 8  
 T37806  
 probable flavoprotein subunit - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 05-May-2000  
 C:Accession: T37806  
 R:Oliver, K.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
 submitted to the EMBL Data Library, September 1997  
 A:Reference number: Z21748  
 A:Accession: T37806  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-513 <OLI>  
 A:Cross-references: EMBL:Z99292; PIDN:CAB16560.1; GSPDB:GN00066; SPDB:SPAC17A2.05

A:Experimental source: strain 972h-; cosmid c17A2  
 C:Genetics:  
 A:Gene: SPDB:SPAC17A2.05  
 A:Map position: 1  
 A:Introns: 71/1  
 C:Superfamily: osmotic growth protein 1; 3-oxosteroid 1-dehydrogenase homology; fumar  
 F:40-330/Domain: fumarate reductase flavoprotein homology <FRF>  
 F:384-487/Domain: 3-oxosteroid 1-dehydrogenase homology <OXD>

Query Match 32.4%; Score 33; DB 2; Length 513;  
 Best Local Similarity 32.0%; Pred. No. 28;  
 Matches 8; Conservative 2; Mismatches 15; Indels 0; Gaps 0;

QY 1 RCLSIKRFXXSXTFIXIXXXMXFF 25  
 ||| | | | | | |  
 DB 2 RCLTIYTWFRRLPPIPTNSASFF 26

RESULT 9  
 T33304  
 hypothetical protein R01B10.5 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T33304  
 R:Langston, Y.; Beck, C.  
 submitted to the EMBL Data Library, May 1998  
 A:Description: The sequence of C. elegans cosmid R01B10.  
 A:Reference number: Z21318  
 A:Accession: T33304  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-670 <LAN>  
 A:Cross-references: EMBL:AF068718; PIDN:AAC17769.1; GSPDB:GN00023; CESP:R01B10.5  
 A:Experimental source: strain Bristol N2; clone R01B10  
 C:Genetics:  
 A:Gene: CESP:R01B10.5  
 A:Map position: 5  
 A:Introns: 30/2; 70/3; 131/3; 181/3; 275/2; 382/2; 430/2; 455/2; 508/3; 545/1; 575/2  
 C:Superfamily: Caenorhabditis elegans hypothetical protein R01B10.5

Query Match 32.4%; Score 33; DB 2; Length 670;  
 Best Local Similarity 35.3%; Pred. No. 35;  
 Matches 6; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 2 CLSIKRFXXSXTFIXI 18  
 | | | | | | | | |  
 DB 606 CFSLTFSGLSSTFLSI 622

RESULT 10  
 T20668  
 hypothetical protein F09F3.7 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Sep-2000  
 C:Accession: T20668  
 R:Thomas, K.  
 submitted to the EMBL Data Library, October 1996  
 A:Reference number: Z19308  
 A:Accession: T20668  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-312 <WIL>  
 A:Cross-references: EMBL:Z81056; PIDN:CAB02905.1; GSPDB:GN00023; CESP:F09F3.7  
 A:Experimental source: clone F09F3  
 C:Genetics:  
 A:Gene: CESP:F09F3.7  
 A:Map position: 5  
 A:Introns: 11/3; 143/1; 198/3  
 C:Superfamily: Caenorhabditis elegans hypothetical protein F32H5.6

Query Match 31.9%; Score 32.5; DB 2; Length 312;  
 Best Local Similarity 32.3%; Pred. No. 23;  
 Matches 10; Conservative 2; Mismatches 18; Indels 1; Gaps 1;



Best Local Similarity 45.5%; Pred. No. 2.1e+02;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 RCLSIKREXXS 11

Db 1332 KCVSVLRFDSS 1342

Search completed: July 16, 2003, 18:09:54  
Job time : 6.7774 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:37:24 ; Search time 1.8887 Seconds  
(without alignments)  
702.728 Million cell updates/sec

Title: US-09-853-079-36

Perfect score: 102

Sequence: 1 RCLSLXRFXXSXXTFIXXXMXFFXXXXXFL 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	36	35.3	376	1 FLHB_TREPA	O83710 treponema p
2	35	34.3	2144	1 CLR2_RAT	Q9gyp2 rattus norv
3	35	34.3	2920	1 CLR2_MOUSE	Q9r0m0 mus musculu
4	35	34.3	2923	1 CLR2_HUMAN	Q9hcu4 homo sapien
5	33	32.4	335	1 CD1D_HUMAN	P15813 homo sapien
6	32	31.4	297	1 Y688_RICPR	Q9zcn5 rickettsia
7	32	31.4	3014	1 CLR1_HUMAN	Q9nyq6 homo sapien
8	32	31.4	3034	1 CLR1_MOUSE	Q35161 mus musculu
9	32	31.4	3301	1 CLR3_MOUSE	Q91zi0 mus musculu
10	32	31.4	3312	1 CLR3_HUMAN	Q9nyq7 homo sapien
11	32	31.4	3313	1 CLR3_RAT	O8B278 rattus norv
12	31	30.4	361	1 EBI2_HUMAN	P32249 homo sapien
13	30	29.4	93	1 UBIL_NPVOP	Q05120 orgyia pseu
14	30	29.4	177	1 RM05_ACACA	P46764 acanthamoeb
15	30	29.4	220	1 TIM2_CHICK	O42146 gallus gall
16	30	29.4	330	1 CPDP_VIBFI	Q56686 vibrio fisc
17	30	29.4	523	1 Y0B4_CAEEL	Q17529 caenorhabdi
18	30	29.4	549	1 X92_TRYBB	P12304 trypanosoma
19	30	29.4	1039	1 YR71_CAEEL	Q09564 caenorhabdi
20	30	29.4	1087	1 YXNK_CLOTRD	P38535 clostridium
21	30	29.4	1142	1 KPC1_NEUCR	P87253 neurospora
22	30	29.4	1640	1 CO3_ONCMY	P98093 oncorhynch
23	29.5	28.9	213	1 YK15_YEAST	P36129 saccharomyc
24	29.5	28.9	648	1 RAD4_SCHPO	P32372 schizosacch
25	29	28.4	229	1 ABME_RAT	P38483 rattus norv
26	29	28.4	363	1 VP43_NPVAC	P34050 autographa
27	29	28.4	395	1 ADMR_MOUSE	P43142 mus musculu
28	29	28.4	395	1 ADMR_RAT	P31392 rattus norv
29	29	28.4	474	1 Y294_MVCGE	Q49411 mycoplasma
30	29	28.4	492	1 NUAM_CHOCR	P48915 chondrus cr
31	29	28.4	498	1 ADRO_SALFO	P82861 salvelinus
32	29	28.4	569	1 ICS1_ARATH	Q957h8 arabidopsis
33	29	28.4	732	1 YMW1_CAEEL	P34489 caenorhabdi

Q92b80 mycoplasma  
O60404 homo sapien  
Q28790 archaeoglob  
P03783 bacterioph  
P47184 saccharomyc  
Q9J5b5 fowlpox vir  
P34577 rattus norv  
P13401 plasmodium  
P51508 mus musculu  
P47855 oryctolagus  
P32250 gallus gall  
O21077 myxine glut

34 28.5 27.9 137 1 Y14A\_MVCGE  
35 28.5 27.9 316 1 OAH3\_HUMAN  
36 28 27.5 85 1 YER2\_ARCFU  
37 28 27.5 112 1 V42\_BPT7  
38 28 27.5 120 1 YJ9V\_YEAST  
39 28 27.5 148 1 V105\_FOWPV  
40 28 27.5 177 1 AG2T\_RAT  
41 28 27.5 215 1 OS25\_PLAGA  
42 28 27.5 229 1 ABME\_MOUSE  
43 28 27.5 236 1 ABME\_RABIT  
44 28 27.5 308 1 P2Y5\_CHICK  
45 28 27.5 318 1 NUIM\_MYXGL

#### ALIGNMENTS

#### RESULT 1

FLHB\_TREPA  
ID FLHB\_TREPA STANDARD; PRT; 376 AA.  
AC O83710:  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Flagellar biosynthetic protein flhb.  
GN FLHB OR TP0715.  
OS Treponema pallidum.  
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
OX NCBI\_TaxID=160;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Nichols;  
RX MEDLINE=98332770; PubMed=9665876;  
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,  
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,  
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
RA Venter J.C.;  
RA "Complete genome sequence of Treponema pallidum, the syphilis  
spirochete.";  
RL Science 281:375-388(1998).  
CC -!- FUNCTION: REQUIRED FOR FORMATION OF THE ROD STRUCTURE IN THE BASAL  
CC BODY OF THE FLAGELLAR APPARATUS. TOGETHER WITH FLII AND FLIH, MAY  
CC CONSTITUTE THE EXPORT APPARATUS OF FLAGELLIN (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
CC (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE FLHB/HRPN/VSCU/SPAS FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE001244; AAC65681.1; -  
CC TIGR; TP0715; -  
CC InterPro; IPR002066; Bac\_export\_2.  
CC Pfam; PF01312; Bac\_export\_2; 1.  
CC PRINTS; PR00950; TYPE31MSPROT.  
CC TIGRFAWS; TIGR00328; flhb; 1.  
KW Flagella; Transmembrane; Inner membrane; Complete proteome.  
FT TRANSMEM 51 71 POTENTIAL.  
FT TRANSMEM 97 117 POTENTIAL.  
FT TRANSMEM 161 181 POTENTIAL.  
FT TRANSMEM 202 222 POTENTIAL.  
SQ SEQUENCE 376 AA; 43039 MW; FE432E2E8B269540 CRC64;

Query Match 35.3%; Score 36; DB 1; Length 376;  
Best Local Similarity 25.8%; Pred. No. 2.3;

CC	EMBL; AB011529; BAA86867.1; -.
DR	HSSP; PF00740; 1EDM.
DR	InterPro; IPRO00152; Asx_hydroxyl.
DR	InterPro; IPRO02126; Cadherin.
DR	InterPro; IPRO00561; EGF-like.
DR	InterPro; IPRO00742; EGF_2.
DR	InterPro; IPRO01881; EGF_Ca.
DR	InterPro; IPRO01438; EGF_II.
DR	InterPro; IPRO00832; GPCR_secretin.
DR	InterPro; IPRO01879; hormn_receptor.
DR	InterPro; IPRO02049; Laminin_EGF.
DR	InterPro; IPRO01791; Laminin_G.
DR	InterPro; IPRO00203; PKD_cys_rich.
DR	Pfam; PF00002; 7tm_2; 1.
DR	Pfam; PF00028; cadherin; 3.
DR	Pfam; PF00008; EGF; 6.
DR	Pfam; PF01825; GPS; 1.
DR	Pfam; PF02793; HRM; 1.
DR	Pfam; PF00054; laminin_G; 1.
DR	PRINTS; PR00205; CADHERIN.
DR	PRINTS; PR00010; EGF_BLOOD.
DR	PRINTS; PR00011; EGF_LAMININ.
DR	PRINTS; PR00249; GPCR_SECRETIN.



FT	DISULFID	1133	1146	BY SIMILARITY.
FT	DISULFID	1149	1161	BY SIMILARITY.
FT	DISULFID	1151	1168	BY SIMILARITY.
FT	DISULFID	1170	1179	BY SIMILARITY.
FT	MOD_RES	816	816	HYDROXYLATION (POTENTIAL).
FT	MOD_RES	1035	1035	HYDROXYLATION (POTENTIAL).
FT	CARBOHYD	261	261	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	301	301	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	407	407	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	726	726	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	790	790	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	966	966	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1052	1052	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1125	1125	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1249	1249	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1268	1268	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1286	1286	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1548	1548	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1570	1570	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	2144 AA; 233480 MW; 6EA898C1BA655ECA CRC64;		
Query Match				
Best Local Similarity		34.3%;	Score 35; DB 1; Length 2144;	
Matches		6; Conservative	2; Mismatches	3; Indels 0; Gaps 0;
QY	1 RCL5XRFXXS 11			
Db	467 RCVSVLRFDSS 477			
RESULT 3				
CLR2_MOUSE				
ID	CLR2_MOUSE	STANDARD;	PRT;	2920 AA.
AC	Q9ROM0; Q922R4; Q99K26;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Flamingo 1)			
DE	(mFm1).			
GN	CELSR2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99418630; PubMed=10490098;			
RA	Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,			
RA	Takeichi M., Uemura T.;			
RT	"Flamingo, a seven-pass transmembrane cadherin, regulates planar cell			
RT	polarity under the control of frizzled.";			
RL	Cell 98:585-595(1999).			
RN	[2]			
RP	SEQUENCE OF 1913-2796 FROM N.A., AND TISSUE SPECIFICITY.			
RX	PubMed=10790539;			
RA	Formstone C.J., Barclay J., Rees M., Little P.F.R.;			
RT	"Chromosomal localization of Celsr2 and Celsr3 in the mouse; Celsr3 is			
RT	a candidate for the tippy (tip) lethal mutant on chromosome 9.";			
RL	Mamm. Genome 11:392-394(2000).			
RN	[3]			
RP	SEQUENCE OF 2014-2920 FROM N.A.			
RC	TISSUE=Breast tumor;			
RA	Straussberg R.;			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	DEVELOPMENTAL STAGE.			
RX	PubMed=11850187;			
RA	Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;			
RT	"Developmental expression profiles of Celsr (Flamingo) genes in the			
RT	mouse.";			
RL	Mech. Dev. 112:157-160(2002).			
CC	-!- FUNCTION: Receptor that may have an important role in cell/cell			

signaling during nervous system formation.

-!- SUBCELLULAR LOCATION: Integral membrane protein.

-!- DEVELOPMENTAL STAGE: Predominantly expressed in the developing CNS, the emerging dorsal root ganglia and cranial ganglia. In the CNS, expression is uniform along the rostrocaudal axis. During gastrulation, it is expressed within the anterior neural ectoderm. At E10, expression is strong in the ventricular zones (VZ) in all sectors of the brain, and lower in the marginal zones (MZ). Between E12 and E15, expression is prominent in the brain. It is strong in VZ, lower in MZ, except in telecephalic MZ where it is predominant. The intensity is higher in all VZ, and lower in differentiating fields than in VZ, except in the cerebral hemispheres, and to a lesser extent in the tectum and cerebellum. A weak expression is also observed in the fetal lungs, kidney and epithelia. In the newborn and postnatal stages, expression remains restricted to the VZ as well as in migrating and postmigratory cells throughout the brain.

-!- TISSUE SPECIFICITY: Expressed in the CNS and in the eye.

-!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

-!- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.

-!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.

-!- SIMILARITY: CONTAINS 1 GPS DOMAIN.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; AB028499; BAA84070.1; -

EMBL; AF031573; AAC68837.1; -

EMBL; BC005499; AAH03499.1; -

DR HSSP; P00740; 1EDM.

DR MGI; 1858235; Celsr2.

DR InterPro; IPR000152; Asx\_hydroxyl.

DR InterPro; IPR002126; Cadherin.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR000742; EGF\_2.

DR InterPro; IPR000832; GPCR\_secretin.

DR InterPro; IPR001879; hormn\_receptor.

DR InterPro; IPR002049; Laminin\_EGF.

DR InterPro; IPR001791; Laminin\_G.

DR InterPro; IPR000203; PKD\_cys\_rich.

DR Pfam; PF00002; 7tm\_2; 1.

DR Pfam; PF00028; cadherin; 9.

DR Pfam; PF00008; EGF; 5.

DR Pfam; PF01825; GPS; 1.

DR Pfam; PF02793; HRM; 1.

DR Pfam; PF00054; laminin\_G; 2.

DR PRINTS; PR00205; CADHERIN.

DR PRINTS; PR00011; EGFLAMININ.

DR PRINTS; PR00249; GPCRSECRETIN.

DR SMART; SM00112; CA; 9.

DR SMART; SM00180; EGF\_Lam; 1.

DR SMART; SM00001; EGF\_like; 6.

DR SMART; SM00303; GPS; 1.

DR SMART; SM00008; Hormr; 1.

DR SMART; SM00282; LamG; 2.

DR PROSITE; PS00010; ASX\_HYDROXYL; 1.

DR PROSITE; PS00232; CADHERIN\_1; 6.

DR PROSITE; PS0268; CADHERIN\_2; 9.

DR PROSITE; PS00022; EGF\_1; 6.

DR PROSITE; PS01186; EGF\_2; 2.

DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.

DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_2; FALSE\_NEG.

DR PROSITE; PS00227; G\_PROTEIN\_RECEP\_F2\_3; 1.

DR PROSITE; PS00261; G\_PROTEIN\_RECEP\_F2\_4; 1.

DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.

DR PROSITE; PS0221; GPS; 1.

DR PROSITE; PS50025; LAM\_G\_DOMAIN; 2;  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
 KW Developmental protein; Hydroxylation; Signal;  
 FT SIGNAL 1 31  
 FT CHAIN 32 2920  
 FT  
 FT DOMAIN 32 2381  
 FT TRANSSEM 2382 2402  
 FT DOMAIN 2403 2414  
 FT TRANSSEM 2415 2434  
 FT DOMAIN 2435 2439  
 FT TRANSSEM 2440 2460  
 FT DOMAIN 2461 2481  
 FT TRANSSEM 2482 2502  
 FT DOMAIN 2503 2519  
 FT TRANSSEM 2520 2540  
 FT DOMAIN 2541 2564  
 FT TRANSSEM 2565 2585  
 FT DOMAIN 2586 2592  
 FT TRANSSEM 2593 2613  
 FT DOMAIN 2614 2920  
 FT DOMAIN 182 289  
 FT DOMAIN 289 399  
 FT DOMAIN 400 506  
 FT DOMAIN 507 611  
 FT DOMAIN 612 713  
 FT DOMAIN 714 816  
 FT DOMAIN 817 922  
 FT DOMAIN 923 1024  
 FT DOMAIN 1029 1147  
 FT DOMAIN 1229 1287  
 FT DOMAIN 1289 1319  
 FT DOMAIN 1329 1367  
 FT DOMAIN 1368 1572  
 FT DOMAIN 1575 1611  
 FT DOMAIN 1615 1792  
 FT DOMAIN 1788 1830  
 FT DOMAIN 1831 1868  
 FT DOMAIN 1884 1923  
 FT DOMAIN 1924 1956  
 FT DOMAIN 2317 2369  
 FT DOMAIN 2744 2749  
 FT DISULFID 1293 1304  
 FT DISULFID 1298 1313  
 FT DISULFID 1315 1318  
 FT DISULFID 1333 1344  
 FT DISULFID 1338 1354  
 FT DISULFID 1356 1366  
 FT DISULFID 1579 1590  
 FT DISULFID 1584 1599  
 FT DISULFID 1601 1610  
 FT DISULFID 1792 1803  
 FT DISULFID 1798 1818  
 FT DISULFID 1820 1829  
 FT DISULFID 1833 1844  
 FT DISULFID 1838 1856  
 FT DISULFID 1858 1867  
 FT DISULFID 1888 1900  
 FT DISULFID 1890 1907  
 FT DISULFID 1909 1922  
 FT DISULFID 1925 1937  
 FT DISULFID 1927 1944  
 FT DISULFID 1946 1955  
 FT MOD\_RES 1592 1592  
 FT CARBOHYD 486 486  
 FT CARBOHYD 558 558  
 FT CARBOHYD 702 702  
 FT CARBOHYD 1037 1037  
 FT CARBOHYD 1077 1077  
 FT CARBOHYD 1183 1183  
 FT CARBOHYD 1213 1213  
 FT CARBOHYD 1502 1502

FT CARBOHYD 1566 1566 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1742 1742 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1828 1828 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1901 1901 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2025 2025 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2062 2062 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2324 2324 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2346 2346 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 2199 2199 V -> L (IN REF. 3).  
 FT CONFLICT 2283 2283 A -> V (IN REF. 3).  
 FT CONFLICT 2535 2535 S -> R (IN REF. 1).  
 FT CONFLICT 2571 2571 L -> R (IN REF. 2).  
 FT CONFLICT 2639 2639 Y -> S (IN REF. 3).  
 Query Match 34.3%; Score 35; DB 1; Length 2920;  
 Best Local Similarity 54.5%; Pred. No. 28;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 RCLSLRFXXS 11  
 Db 1243 RCVSVLRDSS 1253  
 RESULT 4  
 CLR2\_HUMAN  
 ID CLR2\_HUMAN STANDARD; PRT; 2923 AA.  
 AC Q9HCU4; Q92566;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Cadherin EGF LAG seven-pass G-type receptor 2 precursor (Epidermal  
 DE growth factor-like 2) (Multiple epidermal growth factor-like domains  
 DE 3) (Flamingo 1).  
 GN CELSR2 OR CDHF10 OR EGFL2 OR MEGF3 OR KIAA0279.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20363102; PubMed=10907856;  
 RA Vincent J.B., Skaug J., Scherer S.W.;  
 RT "The human homologue of flamingo, EGFL2, encodes a brain-expressed  
 RT large cadherin-like protein with epidermal growth factor-like domains,  
 RT and maps to chromosome 1p13.3-p21.1";  
 RL DNA Res. 7:233-235(2000).  
 RP [2]  
 RP SEQUENCE OF 516-2923 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97191544; PubMed=9039502;  
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,  
 RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;  
 RT "Prediction of the coding sequences of unidentified human genes. VI.  
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
 RT analysis of cDNA clones from cell line KG-1 and brain.";  
 RL DNA Res. 3:321-329(1996).  
 CC -!- FUNCTION: Receptor that may have an important role in cell/cell  
 CC signaling during nervous system formation.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- TISSUE SPECIFICITY: Highest expression in brain and testis.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -!- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@lsb-sib.ch).

CC EMBL; AF234887; AAC00080.1; -  
 CC EMBL; D87469; BAA13407.1; -  
 DR HSP; P15116; INCJ.  
 DR Genew; HGNC:3231; CELSR2.  
 DR MIM; 604265; -  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR000832; GPCR\_secretin.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR000203; PKD\_cys\_rich.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF00028; cadherin; 8.  
 DR Pfam; PF00008; EGF; 6.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR00011; EGF\_LAMININ.  
 DR PRINTS; PR00249; GPCR\_SECRETIN.  
 DR SMART; SM00112; CA; 9.  
 DR SMART; SM00181; EGF; 8.  
 DR SMART; SM00179; EGF\_CA; 5.  
 DR SMART; SM00001; EGF\_like; 7.  
 DR SMART; SM00303; GPS; 1.  
 DR SMART; SM00008; Hormr; 1.  
 DR SMART; SM00282; Lamg; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 2.  
 DR PROSITE; PS00232; CADHERIN\_1; 7.  
 DR PROSITE; PS00268; CADHERIN\_2; 9.  
 DR PROSITE; PS00221; GPS; 1.  
 DR PROSITE; PS00022; EGF\_1; 6.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS00649; G\_PROTEIN\_RECP\_F2\_1; FALSE\_NEG.  
 DR PROSITE; PS00650; G\_PROTEIN\_RECP\_F2\_2; FALSE\_NEG.  
 DR PROSITE; PS00227; G\_PROTEIN\_RECP\_F2\_3; 1.  
 DR PROSITE; PS00261; G\_PROTEIN\_RECP\_F2\_4; 1.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
 KW Developmental protein; Hydroxylation; Signal.  
 FT SIGNAL 1 31  
 FT CHAIN 32 2923  
 FT  
 FT DOMAIN 32 2380  
 FT TRANSMEM 2381 2401  
 FT DOMAIN 2402 2416  
 FT TRANSMEM 2417 2437  
 FT DOMAIN 2438 2438  
 FT TRANSMEM 2439 2459  
 FT DOMAIN 2460 2480  
 FT TRANSMEM 2481 2501  
 FT DOMAIN 2502 2519  
 FT TRANSMEM 2520 2540  
 FT DOMAIN 2541 2560  
 FT TRANSMEM 2561 2581  
 FT DOMAIN 2582 2591  
 FT TRANSMEM 2592 2612  
 FT DOMAIN 2613 2923  
 FT DOMAIN 182 289  
 FT DOMAIN 290 399  
 FT DOMAIN 400 505  
 FT DOMAIN 506 610  
 FT DOMAIN 611 712  
 FT DOMAIN 713 815  
 FT DOMAIN 816 921

FT DOMAIN 922 1023  
 FT DOMAIN 1028 1146  
 FT DOMAIN 1286 1286  
 FT DOMAIN 1288 1324  
 FT DOMAIN 1328 1366  
 FT DOMAIN 1367 1571  
 FT DOMAIN 1574 1610  
 FT DOMAIN 1614 1791  
 FT DOMAIN 1793 1828  
 FT DOMAIN 1829 1867  
 FT DOMAIN 1883 1922  
 FT DOMAIN 1923 1955  
 FT DOMAIN 2316 2368  
 FT DOMAIN 2743 2752  
 FT DISULFID 1232 1243  
 FT DISULFID 1237 1274  
 FT DISULFID 1276 1285  
 FT DISULFID 1292 1303  
 FT DISULFID 1297 1312  
 FT DISULFID 1314 1323  
 FT DISULFID 1332 1343  
 FT DISULFID 1337 1353  
 FT DISULFID 1355 1365  
 FT DISULFID 1578 1589  
 FT DISULFID 1583 1598  
 FT DISULFID 1600 1609  
 FT DISULFID 1797 1808  
 FT DISULFID 1802 1817  
 FT DISULFID 1819 1828  
 FT DISULFID 1832 1843  
 FT DISULFID 1857 1866  
 FT DISULFID 1887 1899  
 FT DISULFID 1889 1906  
 FT DISULFID 1908 1921  
 FT DISULFID 1924 1936  
 FT DISULFID 1926 1943  
 FT DISULFID 1945 1954  
 FT MOD\_RES 1591 1591  
 FT MOD\_RES 1810 1810  
 FT CARBOHYD 486 486  
 FT CARBOHYD 557 557  
 FT CARBOHYD 701 701  
 FT CARBOHYD 1036 1036  
 FT CARBOHYD 1076 1076  
 FT CARBOHYD 1182 1182  
 FT CARBOHYD 1212 1212  
 FT CARBOHYD 1501 1501  
 FT CARBOHYD 1565 1565  
 FT CARBOHYD 1741 1741  
 FT CARBOHYD 1827 1827  
 FT CARBOHYD 1900 1900  
 FT CARBOHYD 2024 2024  
 FT CARBOHYD 2043 2043  
 FT CARBOHYD 2061 2061  
 FT CARBOHYD 2323 2323  
 FT CARBOHYD 2345 2345  
 SQ SEQUENCE 2923 AA; 317447 MW; 382757D315158ED8 CRC64;  
 Query Match 34.3%; Score 35; DB 1; Length 2923;  
 Best Local Similarity 54.5%; Pred. No. 28;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 RCLSLXFFXXS 11  
 ||:|: ||| |  
 Db 1242 RCVSVLRFDS 1252  
 RESULT 5  
 CD1D\_HUMAN  
 ID CD1D\_HUMAN STANDARD; PRT; 335 AA.  
 AC P15813; Q9Y5M4;  
 DT 01-APR-1990 (Rel. 14, Created)



RESULT 7

CLRL\_HUMAN STANDARD; PRT: 3014 AA.

AC Q9NYQ6: Q9Y526; Q9Y522; Q9BWQ5;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Cadherin EGF LAG seven-pass G-type receptor 1 precursor (Flamingo homolog 2) (hFmi2).

DE CELSR1 OR CDHF9 OR FMI2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20202599; PubMed=10716726;

RA Wu Q., Maniatis T.;

RT "Large exons encoding multiple ectodomains are a characteristic feature of protocadherin genes.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=20057165; PubMed=10591208;

RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M., Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K., Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J., Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R., Conway D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E., Dhali P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G., Evans K.L., Fey J.M., Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C., Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A., Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T., Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T., Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L., Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z., Zhan M., Zhang G., Chissole S., Murray J., Miller N., Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T., Schert P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J., Korf I., Bedell J.A., Hillier L., Wardle E., Waterston R., Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyraud M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L., Tillahun Y., Wright H.;

RT "The DNA sequence of human chromosome 22.";

RL Nature 402:489-495(1999).

RN [3]

RP SEQUENCE OF 624-3014 FROM N.A. (ISOFORM 2).

RC TISSUE=Kidney;

RA Strausberg R.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be produced by alternative splicing.

CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

CC -!- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.

CC -!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: AF231024; AAF61930.1; -

CC EMBL: AL021392; CAB50707.1; -

CC EMBL: AL031597; CAB45020.1; ALT\_INIT.

CC EMBL: AL031588; CAB38413.1; -

CC EMBL: BC000059; AAH00059.1; -

CC Genew; HGNC:1850; CELSR1.

CC MIM; 604523; -

CC HSSP; P00749; LURK.

CC InterPro: IPR000152; Asx\_hydroxyl.

CC InterPro: IPR002126; Cadherin.

CC InterPro: IPR000561; EGF-like.

CC InterPro: IPR000832; GPCR\_secretin.

CC InterPro: IPR001879; horrm\_receptor.

CC InterPro: IPR002049; Laminin\_EGF.

CC InterPro: IPR001791; Laminin\_G.

CC InterPro: IPR000203; PKD\_cys\_rich.

CC Pfam: PF000002; 7tm2; 1.

CC Pfam: PF000028; cadherin; 8.

CC Pfam: PF000008; EGF; 6.

CC Pfam: PF01825; GPS; 1.

CC Pfam: PF02793; HRM; 1.

CC Pfam: PF00054; laminin\_G; 1.

CC PRINTS; PR00205; CADHERIN.

CC PRINTS; PR00011; EGF\_LAMININ.

CC PRINTS; PR00249; GPCRSECRETIN.

CC SMART; SM00112; CA; 9.

CC SMART; SM00180; EGF\_Lam; 1.

CC SMART; SM00001; EGF\_like; 6.

CC SMART; SM00303; GPS; 1.

CC SMART; SM00028; Hormr; 1.

CC SMART; SM00282; Lamg; 2.

CC PROSITE; PS00010; ASX\_HYDROXYL; 2.

CC PROSITE; PS00232; CADHERIN\_1; 7.

CC PROSITE; PS00268; CADHERIN\_2; 9.

CC PROSITE; PS00022; EGF\_1; 6.

CC PROSITE; PS01186; EGF\_2; 2.

CC PROSITE; PS0221; GPS; 1.

CC PROSITE; PS00649; G\_PROTEIN\_RECP\_F2\_1; FALSE\_NEG.

CC PROSITE; PS00650; G\_PROTEIN\_RECP\_F2\_2; FALSE\_NEG.

CC PROSITE; PS00227; G\_PROTEIN\_RECP\_F2\_3; 1.

CC PROSITE; PS00261; G\_PROTEIN\_RECP\_F2\_4; 1.

CC PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.

CC PROSITE; PS50025; LAM\_G\_DOMAIN; 2.

CC G-protein coupled receptor; Transmembrane; Glycoprotein; EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat; Developmental protein; Hydroxylation; Signal; Alternative splicing.

FT SIGNAL 1 20

FT CHAIN 21 3014

FT CADHERIN EGF LAG SEVEN-PASS G-TYPE RECEPTOR 1.

FT DOMAIN 22 2469

FT EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 2470 2490

FT DOMAIN 2491 2501

FT CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 2502 2522

FT DOMAIN 2523 2527

FT EXTRACELLULAR (POTENTIAL).

FT	TRANSFEM	2528	2548	3 (POTENTIAL).	DE	Cadherin EGF LAG seven-pass G-type receptor 1 precursor.
FT	DOMAIN	2549	2572	CYTOPLASMIC (POTENTIAL).	GN	CELRL1.
FT	TRANSFEM	2573	2593	4 (POTENTIAL).	OS	Mus musculus (Mouse).
FT	DOMAIN	2594	2611	EXTRACELLULAR (POTENTIAL).	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
FT	TRANSFEM	2612	2632	5 (POTENTIAL).	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
FT	DOMAIN	2633	2655	CYTOPLASMIC (POTENTIAL).	OX	NCBI_TaxID=10090;
FT	TRANSFEM	2656	2676	6 (POTENTIAL).	RN	[1]
FT	DOMAIN	2677	2683	EXTRACELLULAR (POTENTIAL).	RP	SEQUENCE FROM N.A.
FT	TRANSFEM	2684	2704	7 (POTENTIAL).	RX	PubMed=9858697;
FT	DOMAIN	2705	3014	CYTOPLASMIC (POTENTIAL).	RA	Hadjantonakis A.-K., Formstone C.J., Little P.F.R.;
FT	DOMAIN	246	353	CADHERIN 1.	RT	"mCelsrl is an evolutionarily conserved seven-pass transmembrane
FT	DOMAIN	354	459	CADHERIN 2.	RL	receptor and is expressed during mouse embryonic development.";
FT	DOMAIN	460	565	CADHERIN 3.	RN	Mech. Dev. 78:91-95(1998).
FT	DOMAIN	566	687	CADHERIN 4.	RP	TISSUE SPECIFICITY.
FT	DOMAIN	688	789	CADHERIN 5.	RC	STRAIN=C57BL/6; TISSUE=Brain;
FT	DOMAIN	790	892	CADHERIN 6.	RX	MEDLINE=97480720; PubMed=9339365;
FT	DOMAIN	893	999	CADHERIN 7.	RA	Hadjantonakis A.-K., Sheward W.J., Harmor A.J., de Galan L.,
FT	DOMAIN	1000	1101	CADHERIN 8.	RT	Hoovers J.M.N., Little P.F.R.;
FT	DOMAIN	1106	1224	CADHERIN 9.	RT	"Celsrl, a neural-specific gene encoding an unusual seven-pass
FT	DOMAIN	1303	1361	EGF-LIKE 1, CALCIUM-BINDING.	RL	transmembrane receptor, maps to mouse chromosome 15 and human
FT	DOMAIN	1363	1399	EGF-LIKE 2, CALCIUM-BINDING.	RL	chromosome 22qter.";
FT	DOMAIN	1403	1441	EGF-LIKE 3, CALCIUM-BINDING.	RN	Genomics 45:97-104(1997).
FT	DOMAIN	1442	1645	LAMININ G-LIKE 1.	RP	DEVELOPMENTAL STAGE.
FT	DOMAIN	1649	1685	EGF-LIKE 4, CALCIUM-BINDING.	RX	PubMed=11850187;
FT	DOMAIN	1689	1870	LAMININ G-LIKE 2.	RA	Tissir F., De-Backer O., Goffinet A.M., Lambert de Rouvroit C.A.;
FT	DOMAIN	1872	1907	EGF-LIKE 5, CALCIUM-BINDING.	RT	"Developmental expression profiles of Celsr (Flamingo) genes in the
FT	DOMAIN	1908	1946	EGF-LIKE 6, CALCIUM-BINDING.	RT	mouse.";
FT	DOMAIN	1947	1979	EGF-LIKE 7, CALCIUM-BINDING.	RL	Mech. Dev. 112:157-160(2002).
FT	DOMAIN	1981	2016	EGF-LIKE 8, CALCIUM-BINDING.	CC	-!- FUNCTION: Receptor that may have an important role in cell/cell
FT	DOMAIN	2022	2055	LAMININ EGF-LIKE.	CC	-!- SIGNALING: Receptor that may have an important role in cell/cell
FT	DOMAIN	2408	2460	GPS.	CC	-!- SUBCELLULAR LOCATION: Integral membrane protein.
FT	DOMAIN	2659	2663	POLY-LEU.	CC	-!- DEVELOPMENTAL STAGE: First detected at E6. Predominantly expressed
FT	DISULFID	1307	1318	BY SIMILARITY.	CC	in the developing CNS, the emerging dorsal root ganglia and
FT	DISULFID	1312	1349	BY SIMILARITY.	CC	cranial ganglia. In the CNS, expression is uniform along the
FT	DISULFID	1351	1360	BY SIMILARITY.	CC	rostrocaudal axis. During gastrulation, it is expressed in the
FT	DISULFID	1367	1378	BY SIMILARITY.	CC	vicinity of the primitive streak, and becomes predominant in that
FT	DISULFID	1372	1387	BY SIMILARITY.	CC	area at late gastrulation. At E10, detected in ventricular zones
FT	DISULFID	1389	1398	BY SIMILARITY.	CC	(VZ), but not in marginal zones (MZ), and weakly in other
FT	DISULFID	1407	1418	BY SIMILARITY.	CC	structures. Between E12 and E15, a high expression is present in
FT	DISULFID	1412	1428	BY SIMILARITY.	CC	the VZ in all brain areas. No expression in differentiated
FT	DISULFID	1430	1440	BY SIMILARITY.	CC	neural fields. In the newborn and postnatal stages, expression
FT	DISULFID	1653	1664	BY SIMILARITY.	CC	remains restricted to the VZ. Also found weakly in fetal lungs,
FT	DISULFID	1658	1673	BY SIMILARITY.	CC	kidney and epithelia.
FT	DISULFID	1675	1684	BY SIMILARITY.	CC	-!- TISSUE SPECIFICITY: Expressed in the brain, where it is localized
FT	DISULFID	1876	1886	BY SIMILARITY.	CC	principally in the ependymal cell layer, choroid plexus and the
FT	DISULFID	1881	1897	BY SIMILARITY.	CC	area postrema. Also found in spinal chord and in the eye.
FT	DISULFID	1898	1907	BY SIMILARITY.	CC	-!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
FT	DISULFID	1911	1922	BY SIMILARITY.	CC	-!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
FT	DISULFID	1916	1934	BY SIMILARITY.	CC	-!- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.
FT	DISULFID	1936	1945	BY SIMILARITY.	CC	-!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.
FT	DISULFID	1945	1963	BY SIMILARITY.	CC	-!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
FT	DISULFID	1953	1966	BY SIMILARITY.	CC	-!- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAIN.
FT	DISULFID	1968	1978	BY SIMILARITY.	CC	-!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
FT	DISULFID	1985	2000	BY SIMILARITY.	CC	-----
FT	DISULFID	1987	2003	BY SIMILARITY.	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
FT	DISULFID	2005	2015	BY SIMILARITY.	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
FT	DISULFID	2005	2015	BY SIMILARITY.	CC	the European Bioinformatics Institute. There are no restrictions on its
FT	DISULFID	2005	2015	BY SIMILARITY.	CC	use by non-profit institutions as long as its content is in no way
FT	DISULFID	2005	2015	BY SIMILARITY.	CC	modified and this statement is not removed. Usage by and for commercial
FT	DISULFID	2005	2015	BY SIMILARITY.	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
FT	DISULFID	2005	2015	BY SIMILARITY.	CC	or send an email to license@isb-sib.ch).
FT	DISULFID	2005	2015	BY SIMILARITY.	CC	-----
FT	DISULFID	2005	2015	BY SIMILARITY.	CC	EMBL; AF031572; AAC68836.1; -
FT	DISULFID	2005	2015	BY SIMILARITY.	DR	MGD; MGI:1100883; Celsrl.
FT	DISULFID	2005	2015	BY SIMILARITY.	DR	HSP; P00749; LURK.
FT	DISULFID	2005	2015	BY SIMILARITY.	DR	InterPro; IPR000152; Asx_hydroxyl.
FT	DISULFID	2005	2015	BY SIMILARITY.	DR	InterPro; IPR002126; Cadherin.
FT	DISULFID	2005	2015	BY SIMILARITY.	DR	InterPro; IPR000561; EGF-like.
FT	DISULFID	2005	2015	BY SIMILARITY.	DR	InterPro; IPR000832; GPCR_secretin.
FT	DISULFID	2005	2015	BY SIMILARITY.	DR	InterPro; IPR001879; hormn_receptor.
FT	DISULFID	2005	2015	BY SIMILARITY.	DR	InterPro; IPR002049; Laminin_EGF.
FT	DISULFID	2005	2015	BY SIMILARITY.	DR	InterPro; IPR001791; Laminin_G.

Query Match 31.4%; Score 32; DB 1; Length 3014;

Best Local Similarity 45.5%; Pred. No. 1.3e+02;

Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSIKFFXKS 11

:|:| | |

Db 1317 KCVSLRFDDSS 1327

RESULT 8

CLRL\_MOUSE

ID CLRL\_MOUSE

AC O35161;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

STANDARD; PRT; 3034 AA.

DR InterPro: IPR000203; PKD\_cys\_rich.  
DR Pfam: PF00002; 7tm\_2; 1.  
DR Pfam: PF00028; cadherin; 9.  
DR Pfam: PF00008; EGF; 6.  
DR Pfam: PF01825; GPS; 1.  
DR Pfam: PF02793; HRM; 1.  
DR Pfam: PF00053; laminin\_G; 1.  
DR Pfam: PF00054; laminin\_G; 1.  
DR PRINTS: PR00205; CADHERIN.  
DR PRINTS: PR00011; EGF\_LAMININ.  
DR PRINTS: PR00249; GPCRSECRETIN.  
DR SMART: SM00112; CA; 9.  
DR SMART: SM00180; EGF\_Lam; 1.  
DR SMART: SM00001; EGF\_Like; 6.  
DR SMART: SM00303; GPS; 1.  
DR SMART: SM00008; Hornr; 1.  
DR SMART: SM00282; LamG; 2.  
DR PROSITE: PS00010; ASK\_HYDROXYL; 2.  
DR PROSITE: PS00232; CADHERIN\_1; 7.  
DR PROSITE: PS0268; CADHERIN\_2; 9.  
DR PROSITE: PS00022; EGF\_1; 6.  
DR PROSITE: PS01186; EGF\_2; 2.  
DR PROSITE: PS00221; GPS; 1.  
DR PROSITE: PS00649; G-PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
DR PROSITE: PS00650; G-PROTEIN\_RECEP\_F2\_2; FALSE\_NEG.  
DR PROSITE: PS00227; G-PROTEIN\_RECEP\_F2\_3; 1.  
DR PROSITE: PS00261; G-PROTEIN\_RECEP\_F2\_4; 1.  
DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; 1.  
DR PROSITE: PS00025; LAM\_G\_DOMAIN; 2.  
KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
KW Developmental protein; Calcium-binding; Hydroxylation; Signal.  
FT SIGNAL 1 29  
FT CHAIN 30 3034  
FT DOMAIN 21 2484  
FT TRANSMEM 2485 2505  
FT DOMAIN 2506 2517  
FT TRANSMEM 2517 2537  
FT DOMAIN 2538 2542  
FT TRANSMEM 2543 2563  
FT DOMAIN 2564 2587  
FT TRANSMEM 2588 2608  
FT DOMAIN 2609 2625  
FT TRANSMEM 2626 2646  
FT DOMAIN 2647 2670  
FT TRANSMEM 2671 2691  
FT DOMAIN 2692 2694  
FT TRANSMEM 2695 2715  
FT DOMAIN 2716 3034  
FT DOMAIN 261 368  
FT DOMAIN 369 474  
FT DOMAIN 475 580  
FT DOMAIN 581 702  
FT DOMAIN 703 804  
FT DOMAIN 805 907  
FT DOMAIN 908 1014  
FT DOMAIN 1015 1116  
FT DOMAIN 1121 1239  
FT DOMAIN 1318 1376  
FT DOMAIN 1378 1414  
FT DOMAIN 1418 1456  
FT DOMAIN 1457 1661  
FT DOMAIN 1664 1700  
FT DOMAIN 1704 1885  
FT DOMAIN 1887 1922  
FT DOMAIN 1923 1961  
FT DOMAIN 1962 1994  
FT DOMAIN 1996 2031  
FT DOMAIN 2037 2070  
FT DOMAIN 2423 2475  
FT DOMAIN 2674 2678  
FT DOMAIN 1322 1333  
BY SIMILARITY.

FT DISULFID 1327 1364  
FT DISULFID 1366 1375  
FT DISULFID 1382 1393  
FT DISULFID 1387 1402  
FT DISULFID 1404 1413  
FT DISULFID 1422 1433  
FT DISULFID 1427 1443  
FT DISULFID 1445 1455  
FT DISULFID 1668 1679  
FT DISULFID 1673 1688  
FT DISULFID 1690 1699  
FT DISULFID 1891 1902  
FT DISULFID 1896 1911  
FT DISULFID 1913 1922  
FT DISULFID 1926 1937  
FT DISULFID 1931 1949  
FT DISULFID 1951 1960  
FT DISULFID 1960 1978  
FT DISULFID 1968 1981  
FT DISULFID 1983 1993  
FT DISULFID 2000 2015  
FT DISULFID 2002 2018  
FT DISULFID 2020 2030  
FT MOD\_RES 1681 1681  
FT MOD\_RES 1904 1904  
FT CARBOHYD 236 236  
FT CARBOHYD 561 561  
FT CARBOHYD 649 649  
FT CARBOHYD 793 793  
FT CARBOHYD 1129 1129  
FT CARBOHYD 1154 1154  
FT CARBOHYD 1228 1228  
FT CARBOHYD 1264 1264  
FT CARBOHYD 1274 1274  
FT CARBOHYD 1302 1302  
FT CARBOHYD 1591 1591  
FT CARBOHYD 1638 1638  
FT CARBOHYD 1655 1655  
FT CARBOHYD 1994 1994  
FT CARBOHYD 2118 2118  
FT CARBOHYD 2137 2137  
FT CARBOHYD 2144 2144  
FT CARBOHYD 2155 2155  
FT CARBOHYD 2160 2160  
FT CARBOHYD 2272 2272  
FT CARBOHYD 2430 2430  
FT CARBOHYD 2452 2452  
FT CARBOHYD 2538 2538  
SQ SEQUENCE 3034 AA; 330477 MW; EBF38180AF5ED8A8 CRC64;  
Query Match 31.4%; Score 32; DB 1; Length 3034;  
Best Local Similarity 45.5%; Pred. No. 1.3e+02;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 RCLSLXRFXXS 11  
DB 1332 KCVSVLRDSS 1342  
RESULT 9  
CLR3\_MOUSE  
ID CLR3\_MOUSE STANDARD; PRT; 3301 AA.  
AC 091210; O9PSD0;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor.  
GN CELSR3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]





FT DISULFID 2031 2041 BY SIMILARITY.  
 FT DISULFID 2048 2063 BY SIMILARITY.  
 FT DISULFID 2050 2066 BY SIMILARITY.  
 FT DISULFID 2068 2078 BY SIMILARITY.  
 FT MOD\_RES 1952 1952 HYDROXYLATION (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 838 838 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1173 1173 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1308 1308 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1318 1318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1638 1638 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1702 1702 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1759 1759 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2042 2042 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2166 2166 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2185 2185 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2375 2375 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2465 2465 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2497 2497 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 2713 2713 L -> R (IN REF. 2).  
 FT CONFLICT 3024 3024 R -> P (IN REF. 2).  
 SQ SEQUENCE 3301 AA; 358455 MW; A6B18F2DF7F4DEB6 CRC64;

Query Match 31.4%; Score 32; DB 1; Length 3301;  
 Best Local Similarity 45.5%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSI9EXXS 11  
 !:!:|:|:|  
 Db 1380 KCVSLRFDS 1390

RESULT 10  
 CLR3\_HUMAN STANDARD; PRT; 3312 AA.  
 ID Q9NYQ7; O75092;  
 AC Q9NYQ7; O75092;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cadherin EGF LAG seven-pass G-type receptor 3 precursor (Flamingo homolog 1) (hFml1) (Multiple epidermal growth factor-like domains 2) (Epidermal growth factor-like 1).  
 DE (CDLRS3 OR CDHF11 OR FM11 OR EGFL1 OR MEGF2).  
 GN Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20202599; PubMed=10716736;  
 RA Wu Q., Maniatis T.;  
 RT "Large exons encoding multiple ectodomains are a characteristic feature of protocadherin genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).  
 RN [2]  
 RP SEQUENCE OF 1954-3312 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98360089; PubMed=9693030;  
 RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;  
 RT "Identification of high-molecular-weight proteins with multiple EGF-like motifs by motif-trap screening.";  
 RL Genomics 51:27-34(1998).  
 CC -1- FUNCTION: Receptor that may have an important role in cell/cell signaling during nervous system formation.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -1- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.  
 -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; AF231023; AAF61929.1; .  
 CC EMBL; AB011536; BAA32464.1; .  
 CC HSSP; P00740; 1EDM  
 CC Genew; HGNC:3230; CELSR3.  
 CC MTM; 604264; .  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR000832; GPCR\_receptor.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR000203; PKD\_cys\_rich.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF00028; cadherin; 9.  
 DR Pfam; PF00008; EGF; 6.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00034; laminin\_G; 2.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR SMART; SM00112; CA; 8.  
 DR SMART; SM00180; EGF\_Lam; 1.  
 DR SMART; SM00001; EGF\_like; 6.  
 DR SMART; SM00303; GPS; 1.  
 DR SMART; SM00008; HORMR; 1.  
 DR SMART; SM00282; LamG; 2.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; 1.  
 DR PROSITE; PS00232; CADHERIN\_1; 7.  
 DR PROSITE; PS00268; CADHERIN\_2; 8.  
 DR PROSITE; PS00022; EGF\_1; 6.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS00221; GPS; 1.  
 DR PROSITE; PS00649; G\_PROTEIN\_RECEP\_F2\_1; FALSE\_NEG.  
 DR PROSITE; PS00650; G\_PROTEIN\_RECEP\_F2\_2; FALSE\_NEG.  
 DR PROSITE; PS00227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE; PS00261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 DR PROSITE; PS00025; LAM\_G\_DOMAIN; 2.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW EGF-like domain; Calcium-binding; Laminin EGF-like domain; Repeat;  
 KW Developmental protein; Hydroxylation; Signal.  
 FT SIGNAL 1 32  
 FT CHAIN 33 3312  
 FT CADHERIN EGF LAG SEVEN-PASS G-TYPE RECEPTOR 3.  
 FT DOMAIN 33 2540  
 FT TRANSMEM 2541 2561  
 FT DOMAIN 2562 2572  
 FT TRANSMEM 2573 2593  
 FT DOMAIN 2594 2601  
 FT TRANSMEM 2602 2622  
 FT DOMAIN 2623 2643  
 FT TRANSMEM 2644 2664  
 FT DOMAIN 2665 2681  
 FT TRANSMEM 2682 2702  
 FT DOMAIN 2703 2725  
 FT TRANSMEM 2726 2746  
 FT DOMAIN 2747 2753  
 FT TRANSMEM 2754 2774  
 FT DOMAIN 2775 3312  
 FT TRANSMEM 326 433  
 FT DOMAIN 434 545  
 FT CADHERIN 2.

FT	DOMAIN	546	651	CADHERIN 3.	CADHERIN 3.	CLR3_RAT	STANDARD;	PRT;	3313 AA.
FT	DOMAIN	652	756	CADHERIN 4.	CADHERIN 4.	088278;			
FT	DOMAIN	757	858	CADHERIN 5.	CADHERIN 5.	15-JUN-2002 (Rel. 41, Created)			
FT	DOMAIN	859	961	CADHERIN 6.	CADHERIN 6.	15-JUN-2002 (Rel. 41, Last sequence update)			
FT	DOMAIN	962	1067	CADHERIN 7.	CADHERIN 7.	15-JUN-2002 (Rel. 41, Last annotation update)			
FT	DOMAIN	1068	1169	CADHERIN 8.	CADHERIN 8.	Cadherin-EGF LAG seven-pass G-type receptor 3 precursor (Multiple			
FT	DOMAIN	1170	1265	CADHERIN 9.	CADHERIN 9.	epidermal growth factor-like domains 2).			
FT	DOMAIN	1375	1433	EGF-LIKE 1, CALCIUM-BINDING.	EGF-LIKE 1, CALCIUM-BINDING.	CELRS3 OR MEGF2.			
FT	DOMAIN	1435	1471	EGF-LIKE 2, CALCIUM-BINDING.	EGF-LIKE 2, CALCIUM-BINDING.	Rattus norvegicus (Rat).			
FT	DOMAIN	1475	1514	EGF-LIKE 3, CALCIUM-BINDING.	EGF-LIKE 3, CALCIUM-BINDING.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
FT	DOMAIN	1515	1719	LAMININ G-LIKE 1.	LAMININ G-LIKE 1.	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
FT	DOMAIN	1722	1758	EGF-LIKE 4, CALCIUM-BINDING.	EGF-LIKE 4, CALCIUM-BINDING.	NCBI_TaxID=10116;			
FT	DOMAIN	1764	1944	LAMININ G-LIKE 2.	LAMININ G-LIKE 2.	[1]			
FT	DOMAIN	1946	1982	EGF-LIKE 5, CALCIUM-BINDING.	EGF-LIKE 5, CALCIUM-BINDING.	SEQUENCE FROM N.A.			
FT	DOMAIN	1983	2020	EGF-LIKE 6, CALCIUM-BINDING.	EGF-LIKE 6, CALCIUM-BINDING.	STRAIN=Sprague-Dawley; TISSUE=Brain;			
FT	DOMAIN	2021	2053	EGF-LIKE 7, CALCIUM-BINDING.	EGF-LIKE 7, CALCIUM-BINDING.	MEDLINE=98360089; PubMed=9693030;			
FT	DOMAIN	2055	2090	EGF-LIKE 8, CALCIUM-BINDING.	EGF-LIKE 8, CALCIUM-BINDING.	Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;			
FT	DOMAIN	2096	2131	LAMININ EGF-LIKE.	LAMININ EGF-LIKE.	"Identification of high-molecular-weight proteins with multiple			
FT	DOMAIN	2477	2529	GPS.	GPS.	EGF-like motifs by motif-trap screening.";			
FT	DOMAIN	1379	1399	BY SIMILARITY.	BY SIMILARITY.	Genomics 51:27-34(1998).			
FT	DOMAIN	1384	1421	BY SIMILARITY.	BY SIMILARITY.	-!- FUNCTION: Receptor that may have an important role in cell/cell			
FT	DOMAIN	1423	1432	BY SIMILARITY.	BY SIMILARITY.	signaling during nervous system formation.			
FT	DOMAIN	1439	1450	BY SIMILARITY.	BY SIMILARITY.	-!- SUBCELLULAR LOCATION: Integral membrane protein.			
FT	DOMAIN	1444	1459	BY SIMILARITY.	BY SIMILARITY.	-!- TISSUE SPECIFICITY: Expressed in the brain. Expressed in			
FT	DOMAIN	1461	1470	BY SIMILARITY.	BY SIMILARITY.	cerebellum, olfactory bulb, cerebral cortex, hippocampus and			
FT	DOMAIN	1479	1490	BY SIMILARITY.	BY SIMILARITY.	brain stem.			
FT	DOMAIN	1484	1500	BY SIMILARITY.	BY SIMILARITY.	-!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.			
FT	DOMAIN	1502	1513	BY SIMILARITY.	BY SIMILARITY.	-!- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.			
FT	DOMAIN	1726	1737	BY SIMILARITY.	BY SIMILARITY.	-!- SIMILARITY: CONTAINS 8 EGF-LIKE DOMAINS.			
FT	DOMAIN	1731	1746	BY SIMILARITY.	BY SIMILARITY.	-!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.			
FT	DOMAIN	1748	1757	BY SIMILARITY.	BY SIMILARITY.	-!- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.			
FT	DOMAIN	1950	1961	BY SIMILARITY.	BY SIMILARITY.	-!- SIMILARITY: CONTAINS 1 GPS DOMAIN.			
FT	DOMAIN	1955	1970	BY SIMILARITY.	BY SIMILARITY.	-----			
FT	DOMAIN	1972	1981	BY SIMILARITY.	BY SIMILARITY.	This SWISS-PROT entry is copyright. It is produced through a collaboration			
FT	DOMAIN	1985	1996	BY SIMILARITY.	BY SIMILARITY.	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
FT	DOMAIN	1990	2008	BY SIMILARITY.	BY SIMILARITY.	the European Bioinformatics Institute. There are no restrictions on its			
FT	DOMAIN	2010	2019	BY SIMILARITY.	BY SIMILARITY.	use by non-profit institutions as long as its content is in no way			
FT	DOMAIN	2019	2037	BY SIMILARITY.	BY SIMILARITY.	modified and this statement is not removed. Usage by and for commercial			
FT	DOMAIN	2027	2040	BY SIMILARITY.	BY SIMILARITY.	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
FT	DOMAIN	2042	2052	BY SIMILARITY.	BY SIMILARITY.	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
FT	DOMAIN	2059	2074	BY SIMILARITY.	BY SIMILARITY.	-----			
FT	DOMAIN	2061	2077	BY SIMILARITY.	BY SIMILARITY.	EMBL; AB011528; BAA32459.1; -			
FT	DOMAIN	2079	2089	BY SIMILARITY.	BY SIMILARITY.	HSP; P00740; IEDM.			
FT	MOD_RES	1963	1963	HYDROXYLATION (POTENTIAL).	HYDROXYLATION (POTENTIAL).	InterPro: IPR000152; Asx_hydroxyl.			
FT	CARBOHYD	632	632	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	InterPro: IPR002126; Cadherin.			
FT	CARBOHYD	847	847	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	InterPro: IPR000561; EGF-like.			
FT	CARBOHYD	1182	1182	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	InterPro: IPR000742; EGF_2.			
FT	CARBOHYD	1222	1222	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	InterPro: IPR001881; EGF_Ca.			
FT	CARBOHYD	1317	1317	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	InterPro: IPR000832; GPCR_secretin.			
FT	CARBOHYD	1327	1327	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	InterPro: IPR001879; hormn_receptor.			
FT	CARBOHYD	1649	1649	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	InterPro: IPR002049; Laminin_EGF.			
FT	CARBOHYD	1713	1713	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	InterPro: IPR001791; Laminin_G.			
FT	CARBOHYD	1770	1770	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	InterPro: IPR000203; PKD_cys_rich.			
FT	CARBOHYD	2053	2053	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	Pfam: PF00002; 7tm_2; 1.			
FT	CARBOHYD	2177	2177	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	Pfam: PF00028; cadherin; 9.			
FT	CARBOHYD	2196	2196	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	Pfam: PF00008; EGF; 6.			
FT	CARBOHYD	2386	2386	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	Pfam: PF01825; GPS; 1.			
FT	CARBOHYD	2474	2474	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	Pfam: PF02793; HRM; 1.			
FT	CARBOHYD	2506	2506	N-LINKED (GLCNAC. . .) (POTENTIAL).	N-LINKED (GLCNAC. . .) (POTENTIAL).	Pfam: PF00054; laminin_G; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PRINTS; PR00205; CADHERIN.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00112; CA; 8.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00180; EGF_Lam; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00001; EGF_like; 6..			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00303; GPS; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00008; Hormr; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00282; LamG; 2.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00010; ASX_HYDROXYL; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00232; CADHERIN_1; 7.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00268; CADHERIN_2; 8.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00022; EGF_1; 6.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS01186; EGF_2; 4.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00221; GPS; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS50025; LAW_G_DOMAIN; 2.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	-----			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	EMBL; AB011528; BAA32459.1; -			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	HSP; P00740; IEDM.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000152; Asx_hydroxyl.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR002126; Cadherin.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000561; EGF-like.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000742; EGF_2.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR001881; EGF_Ca.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000832; GPCR_secretin.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR001879; hormn_receptor.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR002049; Laminin_EGF.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR001791; Laminin_G.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000203; PKD_cys_rich.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	Pfam: PF00002; 7tm_2; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	Pfam: PF00028; cadherin; 9.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	Pfam: PF00008; EGF; 6.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	Pfam: PF01825; GPS; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	Pfam: PF02793; HRM; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	Pfam: PF00054; laminin_G; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PRINTS; PR00205; CADHERIN.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00112; CA; 8.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00180; EGF_Lam; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00001; EGF_like; 6..			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00303; GPS; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00008; Hormr; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	SMART; SM00282; LamG; 2.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00010; ASX_HYDROXYL; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00232; CADHERIN_1; 7.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00268; CADHERIN_2; 8.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00022; EGF_1; 6.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS01186; EGF_2; 4.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS00221; GPS; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	PROSITE; PS50025; LAW_G_DOMAIN; 2.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	-----			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	EMBL; AB011528; BAA32459.1; -			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	HSP; P00740; IEDM.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000152; Asx_hydroxyl.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR002126; Cadherin.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000561; EGF-like.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000742; EGF_2.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR001881; EGF_Ca.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000832; GPCR_secretin.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR001879; hormn_receptor.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR002049; Laminin_EGF.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR001791; Laminin_G.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	InterPro: IPR000203; PKD_cys_rich.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	Pfam: PF00002; 7tm_2; 1.			
FT	CONFLICT	2158	2158	G -> GLRGAG (IN REF. 2).	G -> GLRGAG (IN REF. 2).	Pfam: PF00028; cadherin; 9.			
FT	CONFLICT	2158</							



```

FT DOMAIN 58 77 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 78 95 2 (POTENTIAL).
FT DOMAIN 96 105 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 106 127 3 (POTENTIAL).
FT DOMAIN 128 149 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 150 168 4 (POTENTIAL).
FT DOMAIN 169 192 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 193 215 5 (POTENTIAL).
FT DOMAIN 216 241 6 (POTENTIAL).
FT TRANSMEM 242 265 6 (POTENTIAL).
FT DOMAIN 266 287 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 288 312 7 (POTENTIAL).
FT DOMAIN 313 361 CYTOPLASMIC (POTENTIAL).
FT DISULFID 104 181 INTERACTION WITH G PROTEINS.
SQ SEQUENCE 361 AA; 41224 MW; B5A2171F34C9C67B CRC64;

Query Match 30.4%; Score 31; DB 1; Length 361;
Best Local Similarity 85.7%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CLSIXRF 8
DB 124 CLSIDRF 130

RESULT 13
ID UBIL_NPVOP STANDARD; PRT; 93 AA.
AC Q05120;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Ubiquitin-like protein.
GN V-UBI.
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMPNV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93286576; PubMed=8389803;
RA Russell R.L.Q., Rohmann G.F.;
RT "Nucleotide sequence of the ubiquitin-39K gene region from the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome.";
RL J. Gen. Virol. 74:1191-1195(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
CC -1- FUNCTION: UBIQUITIN MAY PLAY A ROLE IN VIRAL LIFE CYCLES, OR IN VIRUS-HOST INTERACTIONS. IT IS COVALENTLY LINKED TO COAT PROTEIN SUBUNITS OF SEVERAL DIFFERENT PLANT AND ANIMAL VIRUSES.
CC -1- SIMILARITY: BELONGS TO THE UBIQUITIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13375; BAA02639.1; -
DR EMBL; U75930; AAC59024.1; -
DR PIR; JQ2029; JQ2029.
DR HSSP; P02248; IUBI.
DR InterPro; IPR000626; Ubiquitin.
DR Pfam; PF00240; ubiquitin; 1.

```

```

DR PRINTS; PR00348; UBIQUITIN.
DR SMART; SM00213; UBQ; 1.
DR PROSITE; PS00299; UBIQUITIN_1; 1.
DR PROSITE; PS0053; UBIQUITIN_2; 1.
KW Nuclear protein; Late protein.
FT SITE 48 48 NECESSARY FOR BRANCHED-CHAIN
MULTIUBIQUITIN ADDUCTS.
FT BINDING 76 76 CONJUGATION TO ACCEPTOR PROTEINS.
SQ SEQUENCE 93 AA; 10427 MW; 681B2A99DE964C99 CRC64;

Query Match 29.4%; Score 30; DB 1; Length 93;
Best Local Similarity 62.5%; Pred. No. 12;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RCLSIKRF 8
DB 85 RCLSLIQF 92

RESULT 14
ID RM05_ACACA STANDARD; PRT; 177 AA.
AC P46764;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Mitochondrial 60S ribosomal protein L5.
GN RPL5.
OS Acanthamoeba castellanii (Amoeba).
OG Mitochondrion.
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 30010 / NEFF;
RX MEDLINE=95147275; PubMed=7844823;
RA Burger G., Plante I., Loneragan K.M., Gray M.W.;
RT "The mitochondrial DNA of the amoeboid protozoon, Acanthamoeba castellanii: complete sequence, gene content and genome organization.";
RL J. Mol. Biol. 245:522-537(1995).
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: BELONGS TO THE L5P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12386; AAD11844.1; -
DR InterPro; IPR002132; Ribosomal_L5.
DR Pfam; PF00281; Ribosomal_L5; 1.
DR PROSITE; PS00358; RIBOSOMAL_L5; FALSE_NEG.
KW Ribosomal protein; Mitochondrion.
SQ SEQUENCE 177 AA; 21764 MW; BEA1930DBC375274 CRC64;

Query Match 29.4%; Score 30; DB 1; Length 177;
Best Local Similarity 37.5%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 RCLSIKRFXXSTFI 16
DB 115 KCLKYNRFENSLIYI 130

RESULT 15
ID TIM2_CHICK STANDARD; PRT; 220 AA.
AC O42146;
DT 15-JUL-1998 (Rel. 36, Created)

```

DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Metalloprotease inhibitor 2 precursor (TIMP-2) (Tissue inhibitor of  
DE metalloproteinases-2).  
GN TIMP2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98122528; PubMed=9462696;  
RA Aimes R.T., Li L.H., Weaver B., Hawkes S., Hahn-Dantona E.A.,  
RA Quigley J.P.;  
RT "Cloning, expression, and characterization of chicken tissue  
RT inhibitor of metalloproteinase-2 (TIMP-2) in normal and transformed  
RT chicken embryo fibroblasts.";  
RL J. Cell. Physiol. 174:342-352(1998).  
CC -!- FUNCTION: COMPLEXES WITH METALLOPROTEINASES (SUCH AS COLLAGENASES)  
CC AND IRREVERSIBLY INACTIVATES THEM (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- PTM: THE ACTIVITY OF TIMP-2 IS DEPENDENT ON THE PRESENCE OF  
CC DISULFIDE BONDS.  
CC -!- SIMILARITY: BELONGS TO THE TIMP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF004664; AAB69168.1; -.  
DR HSP: P16035; 2TMP.  
DR InterPro: IPR001820; TIMP.  
DR Pfam: PF00965; TIMP; 1.  
DR SMART: SM00206; TIMP; 1.  
DR PROSITE: PS00288; TIMP; 1.  
KW Metalloprotease inhibitor; Signal.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 220 METALLOPROTEINASE INHIBITOR 2.  
FT DISULFID 27 98 BY SIMILARITY.  
FT DISULFID 29 127 BY SIMILARITY.  
FT DISULFID 39 152 BY SIMILARITY.  
FT DISULFID 154 201 BY SIMILARITY.  
FT DISULFID 159 164 BY SIMILARITY.  
FT DISULFID 172 193 BY SIMILARITY.  
SQ SEQUENCE 220 AA; 24313 MW; 61BDAC760B752E53 CRC64;

Query Match 29.4%; Score 30; DB 1; Length 220;  
Best Local Similarity 63.6%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 RCLISXRFXXS 11  
Db 158 RCLSIPICFVSS 168  
||||| |

Search completed: July 16, 2003, 17:47:26  
Job time : 3.8887 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:37:48 ; Search time 7.77066 Seconds  
(without alignments)  
848.513 Million cell updates/sec

Title: US-09-853-079-36  
Perfect score: 102  
Sequence: 1 RCLSIKXREXXSXTFIIXXXMXFFXXXXXFL 32

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriaph.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	75	73.5	206	5 Q9NIM3	Q9nim3 babesia mic
2	36	35.3	234	10 Q949U7	Q949u7 arabidopsis
3	36	35.3	234	10 Q9LFG6	Q9lfg6 arabidopsis
4	36	35.3	366	2 Q56337	Q56337 treponema p
5	35	34.3	776	5 Q19464	Q19464 caenorhabdi
6	35	34.3	2144	11 Q9QYP2	Q9qyp2 rattus norv
7	35	34.3	2408	4 Q92566	Q92566 homo sapien
8	35	34.3	2920	11 Q9R0M0	Q9r0m0 mus musculu
9	35	34.3	2923	4 Q9HCU4	Q9hcu4 homo sapien
10	34	33.3	106	16 Q9KEL2	Q9kel2 bacillus ha
11	34	33.3	402	4 Q96CLO	Q96cl0 homo sapien
12	34	33.3	492	3 Q96UX9	Q96ux9 hebeloma cy
13	34	33.3	700	5 Q9VWG4	Q9vwg4 drosophila
14	33	32.4	393	16 Q9CHC5	Q9chc5 lactococcus
15	33	32.4	481	5 Q9VWK3	Q9vwk3 drosophila
16	33	32.4	513	3 Q13755	Q13755 schizosacch

17	33	32.4	1086	5 Q9N976	Q9n976 leishmania
18	32.5	31.9	312	5 P90821	P90821 caenorhabdi
19	32	31.4	308	8 Q9TB48	Q9tb48 platynereis
20	32	31.4	347	17 Q9HKZ2	Q9hkz2 thermoplasm
21	32	31.4	355	4 Q9Y506	Q9y506 homo sapien
22	32	31.4	373	16 Q99VB1	Q99vb1 staphylococ
23	32	31.4	406	2 Q9RCB5	Q9rcb5 yersinia ps
24	32	31.4	419	3 Q96X22	Q96x22 magnaporthe
25	32	31.4	439	2 Q8RMF0	Q8rmf0 streptococc
26	32	31.4	492	8 Q92ZQ1	Q92zq1 cyanidiosch
27	32	31.4	512	16 Q9CF17	Q9cf17 lactococcus
28	32	31.4	774	4 Q9BWQ5	Q9bwq5 homo sapien
29	32	31.4	877	10 Q9FHI8	Q9fhi8 arabidopsis
30	32	31.4	1840	3 Q9HED6	Q9hed6 neurospora
31	32	31.4	3014	4 Q9NVQ6	Q9nyq6 homo sapien
32	32	31.4	3034	11 Q35161	Q35161 mus musculu
33	32	31.4	3301	11 Q9LZIO	Q9lzi0 mus musculu
34	32	31.4	3312	4 Q9NRQ7	Q9nyq7 homo sapien
35	32	31.4	3313	11 Q88278	Q88278 rattus norv
36	31.5	30.9	148	11 Q9D213	Q9d213 mus musculu
37	31.5	30.9	528	5 P91256	P91256 caenorhabdi
38	31.5	30.9	560	5 Q9U5A7	Q9u5a7 schistosoma
39	31	30.4	141	11 Q9JUV7	Q9jiv7 mus musculu
40	31	30.4	173	5 Q20416	Q20416 caenorhabdi
41	31	30.4	201	16 Q8XHE0	Q8xhe0 clostridium
42	31	30.4	244	4 Q9H6F9	Q9h6f9 homo sapien
43	31	30.4	260	11 Q9D5Z1	Q9d5z1 mus musculu
44	31	30.4	264	16 P73992	P73992 synechocyst
45	31	30.4	312	12 Q67551	Q67551 garlic late

#### ALIGNMENTS

#### RESULT 1

Q9NIM3 PRELIMINARY; PRT; 206 AA.  
ID Q9NIM3  
AC Q9NIM3;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
DE Seroreactive antigen BMN1-17B.  
OS Babesia microti.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
OX NCBI\_TaxID=5868;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MN1;  
RX MEDLINE=20231818; PubMed=10768973;  
RA Lodes M.J., Houghton R.L., Bruinsma E.S., Mohamath R., Reynolds L.D.,  
RA Benson D.R., Krause P.J., Reed S.G., Persing D.H.;  
RT "Serological expression cloning of novel immunoreactive antigens of  
RT Babesia microti.";  
RL Infect. Immun. 68:2783-2790(2000).  
DR EMBL; AF206526; AAF68254.1;  
SQ SEQUENCE 206 AA; 24963 MW; 4287DE5D8FD15C94 CRC64;

Query Match 73.5%; Score 75; DB 5; Length 206;  
Best Local Similarity 53.1%; Pred. No. 7.5e-09;  
Matches 17; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 RCLSIKXREXXSXTFIIXXXMXFFXXXXXFL 32

Db 45 RCLSIIRFYSSISTFILDFVMPFFLFTYFL 76

#### RESULT 2

Q949U7 PRELIMINARY; PRT; 234 AA.  
ID Q949U7  
AC Q949U7;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)

```

RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; ALI32969; CAB86900.1; -.
DR EMBL; AY054638; AAK96829.1; -.
DR EMBL; AY072493; AAL66908.1; -.
DR HSSP; P30044; 1HD2.
DR InterPro; IPR000866; Ahpc-TSA.
DR Pram; PF00578; Ahpc-TSA; 1.
SQ SEQUENCE 234 AA; 24684 MW; 4F66DA63CD15F003 CRC64;

Query Match 35.3%; Score 36; DB 10; Length 234;
Best Local Similarity 43.8%; Pred. No. 4.4;
Matches 7; Conservative 2; Mismatches 7; Indels 0; Gaps

QY 3 LSIXRFXSXXTFFIXI 18
   || | | | | :
Db 5 LSVSRFMSSSATVISV 20

RESULT 4
Q56337 PRELIMINARY; PRT; 366 AA.
ID AC Q56337;
AC Q56337;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE FlhB'.
GN DE GN FLHB'.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxId=160;
[1]
RP SEQUENCE OF 311-366 FROM N.A.
RN STRAIN-NICHOLS;
RC Hardham J.M., Frye J.G., Young N.R., Stamm L.V.;
RT "Sequences of the flhA, flhF, and orf304 genes of Treponema pallidum.";
RL Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
[2]
RN SEQUENCE OF 1-312 FROM N.A.
RN STRAIN-NICHOLS;
RX MEDLINE=96105201; PubMed=8529894;
RA Hardham J.M., Frye J.G., Stamm L.V.;
RT "Identification and sequences of the Treponema pallidum flm', flhF, flhP, flhQ, flhR and flhB' genes.";
RL Gene 166:57-64(1995).
RL EMBL; U36839; AAB00549.1; -.
DR InterPro; IPR002066; Bac_export_2.
DR Pram; PF01312; Bac_export_2; 1.
DR PRINTS; PR00950; TYPE3IMSPTOT.
DR TIGRFAMS; TIGR00328; flhB; 1.
SQ SEQUENCE 366 AA; 41733 MW; 68DE94348679FF9A CRC64;

Query Match 35.3%; Score 36; DB 2; Length 366;
Best Local Similarity 25.8%; Pred. No. 6.5;
Matches 8; Conservative 4; Mismatches 19; Indels 0; Gaps

QY 2 CLSIXRFXSXXTFFIXIXXMXFPXXXXXFL 32
   | : | : | | | | | | :
Db 75 CIGVLRFETTRATTASQTONTGWFVFVRYPM 105

RESULT 5
Q19464 PRELIMINARY; PRT; 776 AA.
ID AC Q19464;
AC Q19464;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Fl4Fl1.1 protein.
DE GN Fl4Fl1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiidae; Pelodierinae; Caenorhabditis.
```



OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lightning J.;  
 RL Submitted (Oct-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RL Science 282:2012-2018(1998).  
 DR EMBL; 254307; CAA91092.1; -;  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR001622; K-channel\_pore.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR003091; K\_channel.  
 DR InterPro; IPR003131; K\_tetra.  
 DR InterPro; IPR000836; M-channel\_nlg.  
 DR InterPro; IPR003974; Shaw\_channel.  
 DR Pfam; PF00520; Ion\_trans; 2.  
 DR Pfam; PF02214; K\_tetra; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR PRINTS; PR01498; SHAWCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 SQ SEQUENCE 776 AA; 87162 MW; 96A83E3FCBD25F93 CRC64;

Query Match 34.3%; Score 35; DB 5; Length 776;  
 Best Local Similarity 33.3%; Pred. No. 21;  
 Matches 8; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

QY 2 CLSIXRFXSXXTIFIXXXMKFF 25  
 ||:|:|:|:|:|:|:|:|:|  
 DB 255 CFSFHKFVRSPLTIIDVISTGAFF 278

## RESULT 6

Q9QYP2  
 ID Q9QYP2 PRELIMINARY; PRT; 2144 AA.  
 AC Q9QYP2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE MEGF3 (Fragment).  
 GN MEGF3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;  
 RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;  
 RT "Identification of high-molecular-weight proteins with multiple EGF-  
 RT like motifs by motif-trap screening."  
 RL Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: CONTAINS 4 CADHERIN DOMAINS.  
 DR EMBL; AB011529; BAA8687.1; -;  
 DR HSSP; P00740; 1EDM.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001438; EGF\_III.  
 DR InterPro; IPR000832; GPCR\_secretin.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR InterPro; IPR002049; Laminin\_G.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR000203; PKD\_cys\_rich.  
 DR InterPro; IPR001368; TNFR\_c6.  
 DR Pfam; PF00002; 7tm\_2; 1.

DR Pfam; PF00028; cadherin; 3.  
 DR Pfam; PF00008; EGF; 6.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR00010; EGFLOOD.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR SMART; SM00112; CA; 3.  
 DR SMART; SM00179; EGF\_Ca; 1.  
 DR SMART; SM00001; EGF\_like; 6.  
 DR SMART; SM00303; GPS; 1.  
 DR SMART; SM00008; Hormr; 1.  
 DR SMART; SM00282; LamG; 2.  
 DR SMART; SM00208; TNFR; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_2.  
 DR PROSITE; PS00232; CADHERIN\_1; 2.  
 DR PROSITE; PS0268; CADHERIN\_2; 4.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_6.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS0227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE; PS0261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
 KW Calcium-binding; Cell adhesion; EGF-like domain; Glycoprotein.  
 FT NON\_TER 1  
 SQ SEQUENCE 2144 AA; 233480 MW; 6EA898C1BA655ECA CRC64;

Query Match 34.3%; Score 35; DB 11; Length 2144;  
 Best Local Similarity 54.5%; Pred. No. 49;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSIKXRXXS 11  
 ||:|:|:|:|:|:|:|:|:|  
 DB 467 RCVSVLRFDSS 477

## RESULT 7

Q92566  
 ID Q92566 PRELIMINARY; PRT; 2408 AA.  
 AC Q92566;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE KIAA0279 protein (Fragment).  
 GN KIAA0279.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=97191544; PubMed=9039502;  
 RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,  
 RA Tanaka A., Kotani H., Miyajima N., Nomura N.;  
 RT "Prediction of the coding sequences of unidentified human genes. VI.  
 RT the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
 RT analysis of cDNA clones from cell line KG-1 and brain."  
 RL DNA Res. 3:321-329(1996).  
 CC -!- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.  
 DR EMBL; D87469; BAA13407.1; -;  
 DR HSSP; P15116; 1NGJ.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR001438; EGF\_II.  
 DR InterPro; IPR000832; GPCR\_secretin.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR InterPro; IPR002049; Laminin\_G.  
 DR InterPro; IPR001791; Laminin\_G.

DR InterPro; IPR000203; PKD\_cys\_rich.  
 DR InterPro; IPR001368; TNFR\_c6.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF00028; cadherin; 9.  
 DR Pfam; PF00008; EGF; 5.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00054; laminin\_G; 1.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR0010; EGFLOOD.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR SMART; SM00112; CA; 6.  
 DR SMART; SM00179; EGF\_CA; 1.  
 DR SMART; SM00001; EGF\_like; 6.  
 DR SMART; SM00303; GPS; 1.  
 DR SMART; SM00008; Hormr; 1.  
 DR SMART; SM00282; LamG; 2.  
 DR SMART; SM00208; TNFR; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_2.  
 DR PROSITE; PS00232; CADHERIN\_1; 3.  
 DR PROSITE; PS00268; CADHERIN\_2; 6.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_6.  
 DR PROSITE; PS01186; EGF\_2; 4.  
 DR PROSITE; PS00227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE; PS00261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
 KW Calcium-binding; Cell adhesion; EGF-like domain; Glycoprotein.  
 FT NON\_TER 1  
 SQ SEQUENCE 2408 AA; EF4BFC2CF93355F CRC64;

Query Match 34.3%; Score 35; DB 4; Length 2408;  
 Best Local Similarity 54.5%; Pred. No. 54;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 RCLSIKREXXS 11  
 ||:|: || |  
 DB 727 RCVSVLRFDSS 737

RESULT 8  
 Q9ROMO PRELIMINARY; PRT; 2920 AA.  
 AC Q9ROMO  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Flamingo 1.  
 GN CELSR2 OR FLAMINGO 1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99418630; PubMed=10490098;  
 RA Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,  
 RA Takeichi M., Uemura T.;  
 RT "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell  
 RT polarity under the control of frizzled."  
 RL Cell 98:585-595(1999).  
 CC -1- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.  
 DR EMBL; AB028499; BAA84070.1; -;  
 DR HSSP; P00740; IEDM.  
 DR MGD; MGI:1858235; Celser2.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR000832; GPCR\_secretin.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR000203; PKD\_cys\_rich.

DR InterPro; IPR001368; TNFR\_c6.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF00028; cadherin; 9.  
 DR Pfam; PF00008; EGF; 5.  
 DR Pfam; PF01825; GPS; 1.  
 DR Pfam; PF02793; HRM; 1.  
 DR Pfam; PF00054; laminin\_G; 2.  
 DR PRINTS; PR00205; CADHERIN.  
 DR PRINTS; PR00011; EGFLAMININ.  
 DR PRINTS; PR00249; GPCRSECRETIN.  
 DR SMART; SM00112; CA; 9.  
 DR SMART; SM00180; EGF\_Lam; 1.  
 DR SMART; SM00001; EGF\_like; 6.  
 DR SMART; SM00303; GPS; 1.  
 DR SMART; SM00008; Hormr; 1.  
 DR SMART; SM00282; LamG; 2.  
 DR SMART; SM00208; TNFR; 1.  
 DR PROSITE; PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE; PS00232; CADHERIN\_1; 5.  
 DR PROSITE; PS00268; CADHERIN\_2; 9.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_6.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS00227; G\_PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE; PS00261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
 KW Calcium-binding; Cell adhesion; EGF-like domain; Glycoprotein.  
 SQ SEQUENCE 2920 AA; 317648 MW; 2919558DF467114F CRC64;

Query Match 34.3%; Score 35; DB 11; Length 2920;  
 Best Local Similarity 54.5%; Pred. No. 64;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 RCLSIKREXXS 11  
 ||:|: || |  
 DB 1243 RCVSVLRFDSS 1253

RESULT 9  
 Q9HCU4 PRELIMINARY; PRT; 2923 AA.  
 AC Q9HCU4  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE FLAMINGO 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20363102; PubMed=10907856;  
 RA Vincent J.B., Skaug J., Scherer S.W.;  
 RA "The human homologue of flamingo, EGF2, encodes a brain-expressed  
 RT large cadherin-like protein with epidermal growth factor-like domains,  
 RT and maps to chromosome lp13.3-p21.1";  
 RT DNA Res. 7:233-235(2000).  
 CC -1- SIMILARITY: CONTAINS 9 CADHERIN DOMAINS.  
 DR EMBL; AF234887; RAG00080.1; -;  
 DR HSSP; P15116; INCU.  
 DR InterPro; IPR000152; Asx\_hydroxyl.  
 DR InterPro; IPR002126; Cadherin.  
 DR InterPro; IPR000561; EGF-like.  
 DR InterPro; IPR000742; EGF\_2.  
 DR InterPro; IPR001881; EGF\_Ca.  
 DR InterPro; IPR000832; GPCR\_secretin.  
 DR InterPro; IPR001879; hormn\_receptor.  
 DR InterPro; IPR002049; Laminin\_EGF.  
 DR InterPro; IPR001791; Laminin\_G.  
 DR InterPro; IPR000203; PKD\_cys\_rich.  
 DR InterPro; IPR001368; TNFR\_c6.  
 DR Pfam; PF00002; 7tm\_2; 1.  
 DR Pfam; PF00028; cadherin; 8.

DR Pfam: PF00008; EGF; 6.  
 DR Pfam: PF01825; GPS; 1.  
 DR Pfam: PF02793; HRM; 1.  
 DR Pfam: PF00054; laminin\_G; 1.  
 DR PRINTS: PR00205; CADHERIN.  
 DR PRINTS: PR00111; EGF\_LAMININ.  
 DR PRINTS: PR00249; GPCRSECRETIN.  
 DR SMART: SM00112; CA; 9.  
 DR SMART: SM00181; EGF; 8.  
 DR SMART: SM00179; EGF\_CA; 5.  
 DR SMART: SM00001; EGF\_like; 7.  
 DR SMART: SM00303; GPS; 1.  
 DR SMART: SM00008; Hormr; 1.  
 DR SMART: SM00282; LamG; 2.  
 DR SMART: SM00208; TNFR; 1.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE: PS00232; CADHERIN\_1; 6.  
 DR PROSITE: PS02868; CADHERIN\_2; 9.  
 DR PROSITE: PS00022; EGF\_1; UNKNOWN\_6.  
 DR PROSITE: PS01186; EGF\_2; 4.  
 DR PROSITE: PS0227; G-PROTEIN\_RECEP\_F2\_3; 1.  
 DR PROSITE: PS0261; G-PROTEIN\_RECEP\_F2\_4; 1.  
 DR PROSITE: PS01248; LAMININ\_TYPE\_EGF; UNKNOWN\_1.  
 DR KW Calcium-binding; Cell adhesion; Glycoprotein.  
 DR SQ SEQUENCE 2923 AA; 317447 MW; 382757D315158ED8 CRC64;

Query Match 34.3%; Score 35; DB 4; Length 2923;  
 Best Local Similarity 54.5%; Pred. No. 64;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RCLSIREFXXS 11  
 ||: ||: ||  
 Db 1242 RCVSVLRFDSS 1252

RESULT 10  
 Q9KEL2

ID Q9KEL2 PRELIMINARY; PRT; 106 AA.  
 AC Q9KEL2;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Multidrug resistance protein.  
 GN BH0840.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OC NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 halodurans and genomic sequence comparison with Bacillus subtilis."  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL: AP001510; BAB04559.1; -  
 DR InterPro: IPR000390; DUF7.  
 DR InterPro: IPR001092; HLH\_basic.  
 DR Pfam: PF00893; DUF7; 1.  
 DR PROSITE: PS00038; HELIX\_LOOP\_HELIX; UNKNOWN\_1.  
 DR KW Complete proteome.  
 DR SQ SEQUENCE 106 AA; 11537 MW; 23BE07040EGBF303 CRC64;

Query Match 33.3%; Score 34; DB 16; Length 106;  
 Best Local Similarity 42.1%; Pred. No. 6.3;  
 Matches 8; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

QY 7 RFXSXSTFXIXXXMXFF 25  
 :| | ||| ||  
 Db 28 KFIPTVITFVIIASYFF 46

RESULT 11

Q96CLO PRELIMINARY; PRT; 402 AA.  
 ID Q96CLO;  
 AC Q96CLO;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 45.9 kDa protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SKIN;  
 RA Strausberg R.;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC014161; AAH14161.1; -  
 DR InterPro: IPR001909; KRAB.  
 DR InterPro: IPR000822; Znf\_C2H2.  
 DR Pfam: PF00096; zf-C2H2; 11.  
 DR PROSITE: PS08005; KRAB; 1.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; UNKNOWN\_11.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 11.  
 DR KW DNA-binding; Hypothetical protein; Zinc-finger.  
 DR SQ SEQUENCE 402 AA; 45857 MW; 824FBABAF66609C1 CRC64;

Query Match 33.3%; Score 34; DB 4; Length 402;  
 Best Local Similarity 37.5%; Pred. No. 20;  
 Matches 6; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 RCLSIREFXXSXTFI 16  
 ||: ||: |||  
 Db 148 RCIECGFLKHKSTFI 163

RESULT 12

Q96UX9 PRELIMINARY; PRT; 492 AA.  
 ID Q96UX9;  
 AC Q96UX9;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Ammonium transporter.  
 GN AMT3.  
 OS Hebeloma cylindrosporum.  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;  
 OC Agaricales; Cortinariaceae; Hebeloma.  
 OX NCBI\_TaxID=76867;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21460511; PubMed=11576535;  
 RA Javelle A., Rodredquez-Pastrana B.-R., Jacob C., Botton B., Brun A.,  
 RA Andre B., Marini A.-M., Chalot M.;  
 RT "Molecular characterization of two ammonium transporters from the  
 ectomycorrhizal fungus Hebeloma cylindrosporum."  
 RL FEBS Lett. 505:393-398(2001).  
 DR EMBL: AF395543; AAK82417.1; -  
 DR InterPro: IPR001905; Ammonium\_transp.  
 DR Pfam: PF00909; Ammonium\_transp; 1.  
 DR TIGRfam: TIGR00836; amt; 1.  
 DR PROSITE: PS01219; AMMONIUM\_TRANSF; UNKNOWN\_1.  
 DR SQ SEQUENCE 492 AA; 53673 MW; 83EBB1227DFE4064 CRC64;

Query Match 33.3%; Score 34; DB 3; Length 492;  
 Best Local Similarity 30.4%; Pred. No. 23;  
 Matches 7; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY 3 LSIXRFXSXTFXIXXXMXFF 25  
 |: ||: |||  
 Db 4 LATQDFRGRDISFIVAGAMVFF 26

```

RESULT 13
Q9VWG4 PRELIMINARY; PRT; 700 AA.
AC Q9VWG4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG14205 protein.
GN CG14205.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR ENBL: AE003512; AAF48977.1; -.
DR FlyBase; FBgn0031034; CG14205.
DR InterPro; IPR002656; Acyl_transf_3.
DR Pfam; PF01757; Acyl_transf_3; 1.
SQ SEQUENCE 700 AA; 78477 MW; 3514BF1A51A43610 CRC64;

```

```

Query Match 33.3%; Score 34; DB 5; Length 700;
Best Local Similarity 30.4%; Pred. No. 32;
Matches 7; Conservative 4; Mismatches 12; Indels 0; Gaps 0;
Qy 2 CLSIXRFXSXSTXFIIXXXMXF 24
||| : | | : |
Db 151 CLSKLQLASSIPFLAKTAVCF 173

```

```

RESULT 14
Q9CHC5 PRELIMINARY; PRT; 393 AA.
AC Q9CHC5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Unknown protein.
GN Y1BE OR LL0807.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
DR ENBL: AE006314; AAK04905.1; -.
DR InterPro; IPR001117; Cu-oxidase.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 393 AA; 43617 MW; 3BD755460656204C CRC64;

Query Match 32.4%; Score 33; DB 16; Length 393;
Best Local Similarity 26.1%; Pred. No. 32;
Matches 6; Conservative 5; Mismatches 12; Indels 0; Gaps 0;
Qy 3 LSIXRFXSXSTXFIIXXXMXF 25
||| : | | : |
Db 154 ISFERFRASIFLSILVNLIFY 176

RESULT 15
Q9VWK3 PRELIMINARY; PRT; 481 AA.
AC Q9VWK3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG7874 protein.
GN CG7874.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

```

RA Hostin D., Houston K.A., Howland T.J.; Wei M.-H., Ibegwam C.,  
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003511; AAF48935.1; -  
 DR FlyBase; FBgn0030999; CG7874.  
 DR InterPro; IPR002557; Chitin\_bind\_Pera.  
 DR Pfam; PF01607; CBM\_14; 2.  
 DR SMART; SM00494; ChitBD2; 1.  
 SQ SEQUENCE 481 AA; 50912 MW; 4E3F0A7C6241AC4E CRC64;

Query Match 32.4%; Score 33; DB 5; Length 481;  
 Best Local Similarity 42.9%; Pred. No. 38;  
 Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 CLSIXRFXXXTF 15  
 |||: ||  
 Db 342 CLSVGRFAGIDET 355

Search completed: July 16, 2003, 17:49:59  
 Job time : 10.7707 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:36:24 ; Search time 8.95784 Seconds  
(without alignments)  
476.010 Million cell updates/sec

Title: US-09-853-079-39

Perfect score: 108

Sequence: 1 GHXXNKNKXXAXXKSDTQTQEXXXXXXEE 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	75.9	445	19	Babesia microti BM
2	82	75.9	445	20	Babesia microti an
3	82	75.9	445	21	B. microti BMNI-17
4	82	75.9	445	23	Babesia microti an
5	82	75.9	445	23	Babesia microti an
6	82	75.9	445	23	B. microti MN-10/B
7	82	75.9	666	21	Babesia microti an
8	82	75.9	666	23	Babesia microti an
9	82	75.9	677	23	B. microti MN-10/B
10	82	75.9	1132	21	Babesia microti an
	82	75.9	1132	23	Babesia microti an

11	80	74.1	32	20	AAV24359	Babesia microti an
12	80	74.1	32	21	AA30208	B. microti BMNI-17
13	80	74.1	32	23	AB88953	Babesia microti an
14	79	73.1	275	19	AAW56303	Babesia microti BM
15	79	73.1	275	20	AAV24365	Babesia microti an
16	79	73.1	275	21	AA30217	B. microti clone a
17	79	73.1	275	23	AB88962	Babesia microti an
18	66	61.1	25	23	AB89020	Babesia microti an
19	63	58.3	26	23	AB89013	Babesia microti an
20	63	58.3	50	23	AB89012	Babesia microti an
21	54	50.0	25	23	AB89019	Babesia microti an
22	51	47.2	26	23	AB89017	Babesia microti an
23	45	41.7	25	23	AB89016	Babesia microti an
24	44	40.7	25	23	AB89010	Babesia microti an
25	39	36.1	25	23	AB89011	Babesia microti an
26	39	36.1	749	22	ABG05779	Novel human diagno
27	39	36.1	1430	23	AAU98029	S. mutans glucosyl
28	39	36.1	1430	23	AAU98041	S. mutans glucosyl
29	39	36.1	1430	23	AAU98042	S. mutans glucosyl
30	39	36.1	1430	23	AAU98043	S. mutans glucosyl
31	39	36.1	1430	23	AAU98044	S. mutans glucosyl
32	39	36.1	1430	23	AAU98045	S. mutans glucosyl
33	38.5	35.6	1033	22	AB864120	Drosophila melanog
34	38	35.2	25	23	AB89021	Babesia microti an
35	38	35.2	262	22	AA82237	S. epidermidis ope
36	38	35.2	262	23	ABP38901	Staphylococcus epi
37	37	34.3	25	23	AB89015	Babesia microti an
38	37	34.3	367	21	AAV55637	M. genitalium ycfB
39	37	34.3	398	22	ABG03617	Novel human diagno
40	37	34.3	735	23	AAE24135	Human kinase (PKIN
41	36	33.3	400	22	AB868410	Drosophila melanog
42	36	33.3	503	22	AA898980	Murine PCIP1. Mus
43	36	33.3	914	22	AA894420	Human protein sequ
44	35	32.4	407	20	AAV31988	Alpha-ketoglutarat
45	35	32.4	407	22	AA000225	Succinate dehydrog

#### ALIGNMENTS

RESULT 1  
AAW56298  
ID AAW56298 standard; Protein; 445 AA.  
XX  
AC AAW56298;  
XX  
DT 28-SEP-1998 (first entry)  
XX  
DE Babesia microti BMNI-17 complement antigen sequence.  
XX  
KW antigen; detection; diagnosis; vaccine; tick-borne disease;  
KW differentiation; Lyme disease; ehrlichiosis.  
XX  
OS Babesia microti.  
PN EP834567-A2.  
XX  
PD 08-APR-1998.  
XX  
PF 01-OCT-1997; 97EP-0117067.  
XX  
PR 24-APR-1997; 97US-0845258.  
PR 01-OCT-1996; 96US-0723142.  
XX  
XX (CORI-) CORIXA CORP.  
XX Houghton R, Lodes MJ, Reed SG, Sleath PR;  
WPI: 1998-195465/18.  
DR N-PSDB; AAV22753.  
XX  
PT Polypeptides comprising Babesia microti antigens and their  
immunogenic fragments or epitopes - and related nucleic acid,

XX

DT 20-JUN-2002 (first entry)



```
XX Babesia microti antigen epitope #1.
DE Protozoacide; vaccine; antigen; antigenic epitope; infection.
KW Babesia microti.
OS WO200185947-A2.
XX 15-NOV-2001.
XX 09-MAY-2001; 2001WO-US15192.
XX 10-MAY-2000; 2000US-0569098.
XX 27-JUN-2000; 2000US-0605724.
XX 07-SEP-2000; 2000US-0656688.
XX 10-OCT-2000; 2000US-0685436.
XX 13-DEC-2000; 2000US-0737178.
XX 26-FEB-2001; 2001US-0794764.
XX (CORI-) CORIXA CORP.
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;
PI Secrist H;
XX WPI; 2002-216691/27.
XX New Babesia microti antigens, useful for diagnosing and treating B.
PT microti infection, and as component of a composition for enhancing
PT immune response against B. microti infections.
XX Claim 2; Page 191-192; 195pp; English.
XX The present invention relates to novel Babesia microti antigens and their
CC coding sequences. The B. microti antigens, antigenic epitopes of such
CC antigens, and compositions comprising such antigens are useful for
CC diagnosing and treating B. microti infection. The compositions are
CC especially useful for enhancing immune response against B. microti
CC infection. The present sequence was used to illustrate the invention.
XX
XX Sequence 445 AA;
Query Match 75.9%; Score 82; DB 23; Length 445;
Best Local Similarity 56.2%; Pred. No. 5.5e-09;
Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 GHKXNKNKSGNAGIKSYDTQPTQETSDAHEE 32
DB 244 GHDKINKNKGAGIKSYDTQPTQETSDAHEE 275
RESULT 5
ABB89025
ID ABB89025 standard; Protein; 481 AA.
XX AC ABB89025;
XX 20-JUN-2002 (first entry)
XX Babesia microti antigenic epitope fusion protein.
DE DE
XX Protozoacide; vaccine; antigen; antigenic epitope; infection.
XX Babesia microti.
OS WO200185947-A2.
XX 15-NOV-2001.
XX 09-MAY-2001; 2001WO-US15192.
XX 10-MAY-2000; 2000US-0569098.
XX 27-JUN-2000; 2000US-0605724.
XX 07-SEP-2000; 2000US-0656688.
XX 10-OCT-2000; 2000US-0685436.
XX 13-DEC-2000; 2000US-0737178.
XX 26-FEB-2001; 2001US-0794764.
XX (CORI-) CORIXA CORP.
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;
PI Secrist H;
XX WPI; 2002-216691/27.
XX New Babesia microti antigens, useful for diagnosing and treating B.
PT microti infection, and as component of a composition for enhancing
PT immune response against B. microti infections.
XX Example 1; Page 92-93; 195pp; English.
XX The present invention relates to novel Babesia microti antigens and their
CC coding sequences. The B. microti antigens, antigenic epitopes of such
CC antigens, and compositions comprising such antigens are useful for
CC diagnosing and treating B. microti infection. The compositions are
CC especially useful for enhancing immune response against B. microti
CC infection. The present sequence was used to illustrate the invention.
XX
XX Sequence 481 AA;
Query Match 75.9%; Score 82; DB 23; Length 481;
Best Local Similarity 56.2%; Pred. No. 6e-09;
Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1 GHKXNKNKSGNAGIKSYDTQPTQETSDAHEE 32
DB 280 GHDKINKNKGAGIKSYDTQPTQETSDAHEE 311
RESULT 6
AAB30230
ID AAB30230 standard; Protein; 666 AA.
XX AC AAB30230;
XX 12-FEB-2001 (first entry)
XX B. microti MN-10/BMNI-17 fusion protein SEQ ID NO: 85.
DE DE
XX Babesiosis; rodent parasite; tick-borne illness; antigen;
KW disease diagnosis; disease prevention.
XX Babesia sp.
OS Synthetic.
XX WO200060090-A1.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-US09136.
XX 05-APR-1999; 99US-0286488.
XX 17-MAR-2000; 2000US-0528784.
XX (CORI-) CORIXA CORP.
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;
PI WPI; 2000-686939/67.
XX New polypeptides containing an antigenic portion of Babesia microti
PT antigen and DNAs encoding the polypeptides, useful for diagnosing,
PT treating or preventing B. microti infection, or for inducing protective
PT immunity in a patient.
XX Example 7; Page 108-111; 118pp; English.
PS
```

XX The present invention is related to the isolation of antigenic sequences  
 CC from the rodent parasite Babesia microti. This organism is transmitted to  
 CC humans by the same tick which transmits Lyme disease and ehrlichiosis.  
 CC The organism causes a malaria-like infection known as babesiosis. The  
 CC sequences identified by this invention can be used in the diagnosis,  
 CC prevention and treatment of babesiosis.

XX SQ Sequence 666 AA;

Query Match 75.9%; Score 82; DB 21; Length 666;

Best Local Similarity 56.2%; Pred. No. 8.8e-09; Mismatches 14; Indels 0; Gaps 0;

Matches 18; Conservative 0;

QY 1 GHXXNKNKSGNAGIKSYDTQTPQETSDAHEE 496

DB 465 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 496

RESULT 7

ABB88975

ID ABB88975 standard; Protein; 666 AA.

AC ABB88975;

XX 20-JUN-2002 (first entry)

DT Babesia microti antigenic epitope fusion protein BaF-3.

DE Protozoacide; vaccine; antigen; antigenic epitope; infection.

XX Babesia microti.

XX WO200185947-A2.

XX 15-NOV-2001.

XX 09-MAY-2001; 2001WO-US15192.

XX 10-MAY-2000; 2000US-0569098.

XX 27-JUN-2000; 2000US-0605724.

XX 07-SEP-2000; 2000US-0656688.

XX 10-OCT-2000; 2000US-0685436.

XX 13-DEC-2000; 2000US-0737178.

XX 26-FEB-2001; 2001US-0794764.

XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;

XX Secrist H;

XX WPI; 2002-216691/27.

XX New Babesia microti antigens, useful for diagnosing and treating B.

XX microti infection, and as component of a composition for enhancing

XX immune response against B. microti infections

XX Claim 35; Page 113-115; 195pp; English.

XX The present invention relates to novel Babesia microti antigens and their

XX coding sequences. The B. microti antigens, antigenic epitopes of such

XX antigens, and compositions comprising such antigens are useful for

XX diagnosing and treating B. microti infection. The compositions are

XX especially useful for enhancing immune response against B. microti

XX infection. The present sequence was used to illustrate the invention.

XX SQ Sequence 666 AA;

Query Match 75.9%; Score 82; DB 23; Length 666;

Best Local Similarity 56.2%; Pred. No. 8.8e-09;

Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 GHXXNKNKSGNAGIKSYDTQTPQETSDAHEE 496

DB 465 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 496

RESULT 8

ABB88989

ID ABB88989 standard; Protein; 677 AA.

XX ABB88989;

XX 20-JUN-2002 (first entry)

DT Babesia microti antigenic epitope fusion protein BaF-5.

DE Protozoacide; vaccine; antigen; antigenic epitope; infection.

XX Babesia microti.

XX WO200185947-A2.

XX 15-NOV-2001.

XX 09-MAY-2001; 2001WO-US15192.

XX 10-MAY-2000; 2000US-0569098.

XX 27-JUN-2000; 2000US-0605724.

XX 07-SEP-2000; 2000US-0656688.

XX 10-OCT-2000; 2000US-0685436.

XX 13-DEC-2000; 2000US-0737178.

XX 26-FEB-2001; 2001US-0794764.

XX (CORI-) CORIXA CORP.

XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;

XX Secrist H;

XX WPI; 2002-216691/27.

XX New Babesia microti antigens, useful for diagnosing and treating B.

XX microti infection, and as component of a composition for enhancing

XX immune response against B. microti infections

XX Claim 35; Page 160-163; 195pp; English.

XX The present invention relates to novel Babesia microti antigens and their

XX coding sequences. The B. microti antigens, antigenic epitopes of such

XX antigens, and compositions comprising such antigens are useful for

XX diagnosing and treating B. microti infection. The compositions are

XX especially useful for enhancing immune response against B. microti

XX infection. The present sequence was used to illustrate the invention.

XX SQ Sequence 677 AA;

Query Match 75.9%; Score 82; DB 23; Length 677;

Best Local Similarity 56.2%; Pred. No. 8.9e-09;

Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 GHXXNKNKSGNAGIKSYDTQTPQETSDAHEE 496

DB 465 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 496

RESULT 9

AAB30231

ID AAB30231 standard; Protein; 1132 AA.

XX AAB30231;

XX 12-FEB-2001 (first entry)

DT B. microti MN-10/BMNI-17/BMNI-15 fusion protein SEQ ID NO: 87.

XX Babesiosis; rodent parasite; tick-borne illness; antigen;

XX SQ Sequence 677 AA;

Query Match 75.9%; Score 82; DB 23; Length 677;

Best Local Similarity 56.2%; Pred. No. 8.9e-09;

Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1 GHXXNKNKSGNAGIKSYDTQTPQETSDAHEE 496

DB 465 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 496



FT /label= Cys, His  
 PN WO9929869-A1.  
 XX  
 PD 17-JUN-1999.  
 XX  
 PF 11-DEC-1998; 98WO-US26437.  
 XX  
 PR 11-DEC-1997; 97US-0990571.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PA (MAYO-) MAYO FOUNDATION.  
 XX  
 PI Bruinsma E, Houghton R, Lodes MJ, Persing D, Reed SG;  
 PI Sleath PR;  
 XX  
 DR WPI; 1999-385612/32.  
 XX  
 PT New isolated Babesia microti polypeptides  
 XX  
 PS Example 1; Page 94; 126pp; English.  
 XX  
 CC The present invention describes isolated polypeptides comprising  
 CC specific immunogenic portions of Babesia microti. AAX88983 to AAX88994  
 CC encode specifically claimed B. microti immunogenic proteins, and  
 CC AAY24327 to AAY24338 represent these proteins. B. microti polypeptides  
 CC and nucleic acids can be used for detecting B. microti infections. They  
 CC can also be used in vaccines for inducing protective immunity against B.  
 CC microti infections. The present sequence represents a B. microti antigen  
 CC BMNI-17 degenerate repeat sequence.  
 XX  
 SQ Sequence 32 AA;  
 Query Match 74.1%; Score 80; DB 20; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-10;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GHXXNXXNKSXXAXXKSDTQTQXQEXXXXXXEE 32  
 DB 1 GHXXNXXNKSXXAXXKSDTQTQXQEXXXXXXEE 32  
 RESULT 12  
 AAB30208  
 ID AAB30208 standard; Peptide; 32 AA.  
 AC AAB30208;  
 XX  
 DT 12-FEB-2001 (first entry)  
 DE B. microti BMNI-17 antigen reverse complement repeat SEQ ID NO: 39.  
 XX  
 KW Babesiosis; rodent parasite; tick-borne illness; antigen;  
 KW disease diagnosis; disease prevention.  
 XX  
 OS Babesia microti.  
 XX  
 PN WO200060090-A1.  
 XX  
 PD 12-OCT-2000.  
 XX  
 PF 05-APR-2000; 2000WO-US09136.  
 XX  
 PR 05-APR-1999; 99US-0286488.  
 PR 17-MAR-2000; 2000US-0528784.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;  
 PI WPI; 2000-686939/67.  
 XX  
 DR New polypeptides containing an antigenic portion of Babesia microti

PT antigen and DNAs encoding the polypeptides, useful for diagnosing,  
 PT treating or preventing B. microti infection, or for inducing protective  
 XX immunity in a patient  
 PS Claim 6; Page 88; 118pp; English.  
 XX  
 CC The present invention is related to the isolation of antigenic sequences  
 CC from the rodent parasite Babesia microti. This organism is transmitted to  
 CC humans by the same tick which transmits Lyme disease and ehrlichiosis.  
 CC The organism causes a malaria-like infection known as babesiosis. The  
 CC sequences identified by this invention can be used in the diagnosis,  
 CC prevention and treatment of babesiosis.  
 XX  
 SQ Sequence 32 AA;  
 Query Match 74.1%; Score 80; DB 21; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-10;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 GHXXNXXNKSXXAXXKSDTQTQXQEXXXXXXEE 32  
 DB 1 GHXXNXXNKSXXAXXKSDTQTQXQEXXXXXXEE 32  
 RESULT 13  
 ABB88953  
 ID ABB88953 standard; Peptide; 32 AA.  
 AC ABB88953;  
 XX  
 DT 20-JUN-2002 (first entry)  
 DE Babesia microti antigen epitope repeat.  
 XX  
 KW Protozoacide; vaccine; antigen; antigenic epitope; infection.  
 XX  
 OS Babesia microti.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 3 /label= Gly, Asp  
 FT Misc-difference 5 /label= Pro, Ile  
 FT Misc-difference 7 /label= Lys, Thr  
 FT Misc-difference 11 /label= Glu, Gly  
 FT Misc-difference 12 /label= Lys, Asn  
 FT Misc-difference 14 /label= Glu, Gly  
 FT Misc-difference 15 /label= Ile, Arg  
 FT Misc-difference 18 /label= His, Tyr  
 FT Misc-difference 23 /label= Thr, Pro  
 FT Misc-difference 26 /label= Ile, Thr  
 FT Misc-difference 27 /label= Cys, Ser  
 FT Misc-difference 28 /label= Asp, Glu  
 FT Misc-difference 29 /label= Glu, Ala  
 FT Misc-difference 30 /label= Cys, His  
 XX  
 PN WO200185947-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 09-MAY-2001; 2001WO-US15192.

XX 10-MAY-2000; 2000US-0569098.  
 PR 27-JUN-2000; 2000US-0605724.  
 PR 07-SEP-2000; 2000US-0656688.  
 PR 10-OCT-2000; 2000US-0685436.  
 PR 13-DEC-2000; 2000US-0737178.  
 PR 26-FEB-2001; 2001US-0794764.  
 XX (CORI-) CORIXA CORP.  
 XX Read SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;  
 PI Secrist H;  
 XX WPI; 2002-216691/27.  
 XX New Babesia microti antigens, useful for diagnosing and treating B.  
 PT microti infection, and as component of a composition for enhancing  
 PT immune response against B. microti infections -  
 XX Claim 7; Page 94; 195pp; English.  
 XX The present invention relates to novel Babesia microti antigens and their  
 CC coding sequences. The B. microti antigens, antigenic epitopes of such  
 CC antigens, and compositions comprising such antigens are useful for  
 CC diagnosing and treating B. microti infection. The compositions are  
 CC especially useful for enhancing immune response against B. microti  
 CC infection. The present sequence was used to illustrate the invention.  
 XX  
 XX Sequence 32 AA;  
 SQ  
 Query Match 74.1%; Score 80; DB 23; Length 32;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-10;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 GHXXNKNKXXAXXKSDTQTQXQEXXXXXXEE 32  
 |||||  
 Db 1 GHXXNKNKXXAXXKSDTQTQXQEXXXXXXEE 32  
 |||||  
 RESULT 14  
 AAW56303  
 ID AAW56303 standard; Protein; 275 AA.  
 AC AAW56303;  
 XX  
 XX 28-SEP-1998 (first entry)  
 DT  
 XX Babesia microti BMNI-20 antigen sequence.  
 DE antigen; detection; diagnosis; vaccine; tick-borne disease;  
 XX differentiation; Lyme disease; ehrlichiosis.  
 KW Babesia microti.  
 XX EP834567-A2.  
 PN  
 XX 08-APR-1998.  
 PD  
 XX 01-OCT-1997; 97EP-0117067.  
 PF  
 XX 24-APR-1997; 97US-0845258.  
 PR 01-OCT-1996; 96US-0723142.  
 XX (CORI-) CORIXA CORP.  
 XX Houghton R, Lodes MJ, Reed SG, Sleath PR;  
 PI WPI; 1998-195465/18.  
 XX N-PSDB; AAW22749.  
 DR Polypeptides comprising Babesia microti antigens and their  
 XX immunogenic fragments or epitopes - and related nucleic acid,  
 PT vectors, transformed cells and antibodies, useful for diagnosis of  
 PT

PT infection and in protective vaccines  
 XX Claim 1; Page 101-102; 113pp; English.  
 PS  
 XX The sequence is that of a polypeptide comprising at least  
 CC one antigenic portion of a Babesia microti antigen. It can be used  
 CC to diagnose B. microti infection by detecting specific antibodies  
 CC in usual immunoassays. Infection can also be diagnosed using:  
 CC (a) primers or probes derived from the coding sequence, in  
 CC standard amplification or hybridisation tests, or (b) using  
 CC antibodies to detect the corresponding antigen. It is also  
 CC useful in vaccines to protect against infection, especially  
 CC when formulated with an adjuvant. The new diagnostic methods  
 CC allow rapid differentiation between B. microti infection and  
 CC other tick-borne diseases (Lyme disease and ehrlichiosis) that  
 CC have similar symptoms but require different treatments.  
 XX  
 XX Sequence 275 AA;  
 SQ  
 Query Match 73.1%; Score 79; DB 19; Length 275;  
 Best Local Similarity 56.2%; Pred. No. 1.4e-08;  
 Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 Qy 1 GHXXNKNKXXAXXKSDTQTQXQEXXXXXXEE 32  
 |||||  
 Db 216 GHGKPNTKSEKAEKSHDTQTQEQICECEE 247  
 |||||  
 RESULT 15  
 AAY24365  
 ID AAY24365 standard; Protein; 275 AA.  
 XX  
 XX AC AAY24365;  
 XX  
 XX 16-SEP-1999 (first entry)  
 DT Babesia microti antigen BMNI-20 complementary open reading frame protein.  
 XX Babesia microti; antigen; immunogen; diagnosis; infection; vaccine;  
 DE immunity; detection.  
 KW Babesia microti.  
 XX OS WO9929869-A1.  
 PN  
 XX 17-JUN-1999.  
 PD  
 XX 11-DEC-1998; 98WO-US26437;  
 PF  
 XX 11-DEC-1997; 97US-0990571.  
 PR (CORI-) CORIXA CORP.  
 XX (MAYO-) MAYO FOUNDATION.  
 PA Bruinisma E, Houghton R, Lodes MJ, Persing D, Reed SG;  
 XX Sleath PR;  
 PI WPI; 1999-385612/32.  
 XX N-PSDB; AAX90017.  
 DR  
 XX New isolated Babesia microti polypeptides  
 PT  
 XX Example 1; Page 109-110; 126pp; English.  
 PS  
 XX The present invention describes isolated polypeptides comprising  
 CC specific immunogenic portions of Babesia microti. AAX88983 to AAX88994  
 CC encode specifically claimed B. microti immunogenic proteins, and  
 CC AAY24327 to AAY24338 represent these proteins. B. microti polypeptides  
 CC and nucleic acids can be used for detecting B. microti infections. They  
 CC can also be used in vaccines for inducing protective immunity against B.  
 CC microti infections. The present sequence represents a B. microti antigen.  
 XX  
 XX Sequence 275 AA;  
 SQ

Search completed: July 16, 2003, 17:46:44  
Job time : 9.95784 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	82	75.9	445	4	US-08-845-258-38		Sequence 38, Appl
2	82	75.9	445	4	US-08-990-571-38		Sequence 38, Appl
3	82	75.9	445	4	US-08-723-142A-38		Sequence 38, Appl
4	82	75.9	445	4	US-09-528-784A-38		Sequence 38, Appl
5	82	75.9	666	4	US-09-528-784A-85		Sequence 85, Appl
6	82	75.9	1132	4	US-09-528-784A-87		Sequence 87, Appl
7	80	74.1	32	4	US-08-845-258-39		Sequence 39, Appl
8	80	74.1	32	4	US-08-990-571-39		Sequence 39, Appl
9	80	74.1	32	4	US-08-723-142A-39		Sequence 39, Appl
10	80	74.1	32	4	US-09-528-784A-39		Sequence 39, Appl
11	79	73.1	275	4	US-08-845-258-53		Sequence 53, Appl
12	79	73.1	275	4	US-08-990-571-53		Sequence 53, Appl
13	79	73.1	275	4	US-09-528-784A-53		Sequence 53, Appl
14	39	36.1	1430	3	US-09-008-172-2		Sequence 2, Appl
15	39	36.1	1430	4	US-09-210-361-6		Sequence 6, Appl
16	39	36.1	1430	4	US-09-740-274-6		Sequence 6, Appl
17	38	35.2	262	4	US-09-134-001C-3746		Sequence 3746, Ap
18	35	32.4	407	4	US-09-271-438A-4		Sequence 4, Appl
19	35	32.4	407	4	US-09-271-438A-10		Sequence 10, Appl
20	34	31.5	1048	3	US-09-356-952-5		Sequence 5, Appl
21	34	31.5	1141	1	US-08-363-300-2		Sequence 2, Appl
22	33	30.6	3696	4	US-09-134-001C-5080		Sequence 5080, Ap
23	32	29.6	245	2	US-08-897-340-35		Sequence 35, Appl
24	32	29.6	245	4	US-09-252-329-35		Sequence 35, Appl
25	32	29.6	264	1	US-08-482-271-3		Sequence 3, Appl
26	32	29.6	264	1	US-08-482-271-4		Sequence 4, Appl
27	32	29.6	264	2	US-08-854-811-45		Sequence 45, Appl

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723.142A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-723-142A-38
Query Match 75.9%; Score 82; DB 4; Length 445;
Best Local Similarity 56.2%; Pred. No. 1.3e-09;
Matches 18; Conservative 0; Mismatches 14; Indels

QY 1 GHXXNKNKXXXXXSDTQTQXQEXXXXXXEE 32
||| ||| ||| ||| ||| ||| |||
DB 244 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 275

RESULT 4
US-09-528-784A-38
; Sequence 38, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528.784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Babesia microti
;
US-09-528-784A-38
Query Match 75.9%; Score 82; DB 4; Length 445;
Best Local Similarity 56.2%; Pred. No. 1.3e-09;
Matches 18; Conservative 0; Mismatches 14; Indels

QY 1 GHXXNKNKXXXXXSDTQTQXQEXXXXXXEE 32
||| ||| ||| ||| ||| ||| |||
DB 244 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 275

RESULT 5
US-09-528-784A-85
; Sequence 85, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528.784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Babesia microti
;
US-09-528-784A-85
Query Match 75.9%; Score 82; DB 4; Length 445;
Best Local Similarity 56.2%; Pred. No. 1.3e-09;
Matches 18; Conservative 0; Mismatches 14; Indels

QY 1 GHXXNKNKXXXXXSDTQTQXQEXXXXXXEE 32
||| ||| ||| ||| ||| ||| |||
DB 244 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 275

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723.142A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-723-142A-38

Query Match 75.9%; Score 82; DB 4; Length 445;
Best Local Similarity 56.2%; Pred. No. 1.3e-09;
Matches 18; Conservative 0; Mismatches 14; Indels

QY 1 GHXXNKNKXXXXXSDTQTQXQEXXXXXXEE 32
||| ||| ||| ||| ||| ||| |||
DB 244 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 275

RESULT 4
US-09-528-784A-38
; Sequence 38, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528.784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Babesia microti
; US-09-528-784A-38

Query Match 75.9%; Score 82; DB 4; Length 445;
Best Local Similarity 56.2%; Pred. No. 1.3e-09;
Matches 18; Conservative 0; Mismatches 14; Indels

QY 1 GHXXNKNKXXXXXSDTQTQXQEXXXXXXEE 32
||| ||| ||| ||| ||| ||| |||
DB 244 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 275

RESULT 5
US-09-528-784A-85
; Sequence 85, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528.784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Babesia microti
; US-09-528-784A-85

```



```
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 85
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Babesia
US-09-528-784A-85

Query Match      75.9%; Score 82; DB 4; Length 666;
Best Local Similarity 56.2%; Pred. No. 2.le-09;
Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      1 GHXXNXXKXXAXXKSDTQTXQEXXXXXXEE 32
Db      11 | | | | | | | | | | | | | | | |
465 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 496

RESULT 6
US-09-528-784A-87
; Sequence 87, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 87
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Babesia
US-09-528-784A-87

Query Match      75.9%; Score 82; DB 4; Length 1132;
Best Local Similarity 56.2%; Pred. No. 3.8e-09;
Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      1 GHXXNXXKXXAXXKSDTQTXQEXXXXXXEE 32
Db      11 | | | | | | | | | | | | | | | |
931 GHDKINKNKGAGIKSYDTQTPQETSDAHEE 962

RESULT 7
US-08-845-258-39
; Sequence 39, Application US/08845258
; Patent No. 6183976
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,258
; FILING DATE: 24-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "Residue can be either Gly
; OTHER INFORMATION: or Asp"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5
; OTHER INFORMATION: /note= "Residue can be either Pro
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 7
; OTHER INFORMATION: /note= "Residue can be either Lys
; OTHER INFORMATION: or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 11
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: or Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 12
; OTHER INFORMATION: /note= "Residue can be either Lys
; OTHER INFORMATION: or Asn"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: or Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: or Arg"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 18
; OTHER INFORMATION: /note= "Residue can be either His
; OTHER INFORMATION: or Tyr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "Residue can be either Thr
; OTHER INFORMATION: or Pro"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
```

OTHER INFORMATION: /note= "Residue can be either Cys  
OTHER INFORMATION: or Ser"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 28  
OTHER INFORMATION: /note= "Residue can be either Asp  
OTHER INFORMATION: or Glu"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 29  
OTHER INFORMATION: /note= "Residue can be either Glu  
OTHER INFORMATION: or Ala"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 30  
OTHER INFORMATION: /note= "Residue can be either Cys  
OTHER INFORMATION: or His"  
US-08-845-258-39

Query Match 74.1%; Score 80; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 1.9e-10;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHXXNKNKXXAXXKSDTQTQXQEXXXXXXEE 32  
|||||  
Db 1 GHXXNKNKXXAXXKSDTQTQXQEXXXXXXEE 32

RESULT 8  
US-08-990-571-39  
Sequence 39, Application US/08990571  
Patent No. 6214971  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G. et al.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/990,571  
FILING DATE: 11-DEC-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.426C2  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 3  
OTHER INFORMATION: /note= "Residue can be either Gly  
OTHER INFORMATION: or Asp"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 5

OTHER INFORMATION: /note= "Residue can be either Pro  
OTHER INFORMATION: or Ile"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 7  
OTHER INFORMATION: /note= "Residue can be either Lys  
OTHER INFORMATION: or Thr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 11  
OTHER INFORMATION: /note= "Residue can be either Glu  
OTHER INFORMATION: or Gly"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 12  
OTHER INFORMATION: /note= "Residue can be either Lys  
OTHER INFORMATION: or Asn"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 14  
OTHER INFORMATION: /note= "Residue can be either Glu  
OTHER INFORMATION: or Gly"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 15  
OTHER INFORMATION: /note= "Residue can be either Ile  
OTHER INFORMATION: or Arg"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 18  
OTHER INFORMATION: /note= "Residue can be either His  
OTHER INFORMATION: or Tyr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 23  
OTHER INFORMATION: /note= "Residue can be either Thr  
OTHER INFORMATION: or Pro"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 26  
OTHER INFORMATION: /note= "Residue can be either Ile  
OTHER INFORMATION: or Thr"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 27  
OTHER INFORMATION: /note= "Residue can be either Cys  
OTHER INFORMATION: or Ser"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 28  
OTHER INFORMATION: /note= "Residue can be either Asp  
OTHER INFORMATION: or Glu"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 29  
OTHER INFORMATION: /note= "Residue can be either Glu  
OTHER INFORMATION: or Ala"  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 30  
OTHER INFORMATION: /note= "Residue can be either Cys  
OTHER INFORMATION: or His"  
US-08-990-571-39

Query Match 74.1%; Score 80; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 1.9e-10;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHXXNKNKXXAXXKSDTQTQXQEXXXXXXEE 32  
|||||  
Db 1 GHXXNKNKXXAXXKSDTQTQXQEXXXXXXEE 32

```
RESULT 9
US-08-723-142A-39
; Sequence 39, Application US/08723142A
; Patent No. 6306396
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF B. MICROTI INFECTION
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,142A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "Residue can be either Gly
; OTHER INFORMATION: or Asp"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5
; OTHER INFORMATION: /note= "Residue can be either Pro
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 7
; OTHER INFORMATION: /note= "Residue can be either Lys
; OTHER INFORMATION: or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 11
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: or Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 12
; OTHER INFORMATION: /note= "Residue can be either Lys
; OTHER INFORMATION: or Asn"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: or Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: or Arg"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 18
; OTHER INFORMATION: /note= "Residue can be either His
; OTHER INFORMATION: or Tyr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "Residue can be either Thr
; OTHER INFORMATION: or Pro"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: or Ser"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 28
; OTHER INFORMATION: /note= "Residue can be either Asp
; OTHER INFORMATION: or Glu"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 29
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: or Ala"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 30
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: or His"
; US-08-723-142A-39
; Query Match 74.1%; Score 80; DB 4; Length 32;
; Best Local Similarity 100.0%; Pred. No. 1.9e-10;
; Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 GHXXNKNKSNXXKXSDTQTXOEXXXEE 32
; | | | | | | | | | | | | | | | | | |
; DB 1 GHXXNKNKSNXXKXSDTQTXOEXXXEE 32
;
; RESULT 10
US-09-528-784A-39
; Sequence 39, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Babesia microti
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3)...(3)
```

OTHER INFORMATION: Xaa = Glycine or Aspartic Acid  
NAME/KEY: VARIANT  
LOCATION: (5)...(5)  
OTHER INFORMATION: Xaa = Proline or Isoleucine  
NAME/KEY: VARIANT  
LOCATION: (7)...(7)  
OTHER INFORMATION: Xaa = Lysine or Threonine  
NAME/KEY: VARIANT  
LOCATION: (11)...(11)  
OTHER INFORMATION: Xaa = Glutamic Acid or Glycine  
NAME/KEY: VARIANT  
LOCATION: (12)...(12)  
OTHER INFORMATION: Xaa = Lysine or Asparagine  
NAME/KEY: VARIANT  
LOCATION: (14)...(14)  
OTHER INFORMATION: Xaa = Glutamic Acid or Glycine  
NAME/KEY: VARIANT  
LOCATION: (15)...(15)  
OTHER INFORMATION: Xaa = Isoleucine or Arginine  
NAME/KEY: VARIANT  
LOCATION: (18)...(18)  
OTHER INFORMATION: Xaa = Histidine or Tyrosine  
NAME/KEY: VARIANT  
LOCATION: (23)...(23)  
OTHER INFORMATION: Xaa = Threonine or Proline  
NAME/KEY: VARIANT  
LOCATION: (26)...(26)  
OTHER INFORMATION: Xaa = Isoleucine or Threonine  
NAME/KEY: VARIANT  
LOCATION: (27)...(27)  
OTHER INFORMATION: Xaa = Cysteine or Serine  
NAME/KEY: VARIANT  
LOCATION: (28)...(28)  
OTHER INFORMATION: Xaa = Aspartic Acid or Glutamic Acid  
NAME/KEY: VARIANT  
LOCATION: (29)...(29)  
OTHER INFORMATION: Xaa = Glutamic Acid or Alanine  
NAME/KEY: VARIANT  
LOCATION: (30)...(30)  
OTHER INFORMATION: Xaa = Cysteine or Histidine  
US-09-528-784A-39

Query Match 74.1%; Score 80; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 1.9e-10;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHXXNKNKXXAXXKSDTQTQXQXXXXXEE 32  
Db 1 GHXXNKNKXXAXXKSDTQTQXQXXXXXEE 32

RESULT 11  
US-08-845-258-53  
Sequence 53, Application US/08845258  
Patent No. 6183976  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Lodes, Michael J.  
APPLICANT: Houghton, Raymond  
APPLICANT: Sleath, Paul R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
AND TREATMENT OF B. MICROTI INFECTION  
NUMBER OF SEQUENCES: 53  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/845,258  
FILING DATE: 24-APR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.426C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 275 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Babesia Microti  
US-08-845-258-53

Query Match 73.1%; Score 79; DB 4; Length 275;  
Best Local Similarity 56.2%; Pred. No. 3.4e-09;  
Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 1 GHXXNKNKXXAXXKSDTQTQXQXXXXXEE 32  
Db 216 GHGKPNKSKAERKSHDTQTQICECEE 247

RESULT 12  
US-08-990-571-53  
Sequence 53, Application US/08990571  
Patent No. 6214971  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G. et al.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/990,571  
FILING DATE: 11-DEC-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.426C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 275 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Babesia Microti





Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	82	75.9	445	9	US-09-286-488-38	Sequence 38, Appl
2	82	75.9	445	9	US-09-853-079-38	Sequence 38, Appl
3	82	75.9	445	10	US-09-737-178-38	Sequence 38, Appl
4	82	75.9	481	9	US-09-853-079-211	Sequence 211, Appl
5	82	75.9	666	9	US-09-853-079-85	Sequence 85, Appl
6	82	75.9	666	10	US-09-737-178-85	Sequence 85, Appl
7	82	75.9	677	9	US-09-853-079-144	Sequence 144, Appl
8	82	75.9	677	10	US-09-737-178-144	Sequence 144, Appl
9	82	75.9	1132	9	US-09-853-079-87	Sequence 87, Appl
10	82	75.9	1132	10	US-09-737-178-87	Sequence 87, Appl
11	80	74.1	32	9	US-09-286-488-39	Sequence 39, Appl
12	80	74.1	32	9	US-09-853-079-39	Sequence 39, Appl
13	80	74.1	32	10	US-09-737-178-39	Sequence 39, Appl
14	79	73.1	275	9	US-09-286-488-53	Sequence 53, Appl
15	79	73.1	275	9	US-09-853-079-53	Sequence 53, Appl
16	79	73.1	275	10	US-09-737-178-53	Sequence 53, Appl
17	66	61.1	25	9	US-09-853-079-202	Sequence 202, Appl
18	63	58.3	26	9	US-09-853-079-195	Sequence 195, Appl
19	63	58.3	50	9	US-09-853-079-194	Sequence 194, Appl

RESULT 2  
US-09-853-079-38  
; Sequence 38, Application US/09853079.  
; Publication No. US2003010969A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.

1

; SEQ ID NO 85





```
Db      931 GHDKINKSGNAGIKSYDTQTQETSADHEE 962
|| | | | | | | | | | | | | | | | |
RESULT 11
US-09-286-488-39
; Sequence 39, Application US/09286488
; Patent No. US20020169136A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C3
; CURRENT APPLICATION NUMBER: US/09/286,488
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Babesia microti
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3)...(3)
; OTHER INFORMATION: Xaa = Glycine or Aspartic Acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (5)...(5)
; OTHER INFORMATION: Xaa = Proline or Isoleucine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (7)...(7)
; OTHER INFORMATION: Xaa = Lysine or Threonine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (11)...(11)
; OTHER INFORMATION: Xaa = Glutamic Acid or Glycine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (12)...(12)
; OTHER INFORMATION: Xaa = Lysine or Asparagine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (14)...(14)
; OTHER INFORMATION: Xaa = Glutamic Acid or Glycine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (15)...(15)
; OTHER INFORMATION: Xaa = Isoleucine or Arginine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (18)...(18)
; OTHER INFORMATION: Xaa = Histidine or Tyrosine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (23)...(23)
; OTHER INFORMATION: Xaa = Threonine or Proline
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (26)...(26)
; OTHER INFORMATION: Xaa = Isoleucine or Threonine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (27)...(27)
; OTHER INFORMATION: Xaa = Cysteine or Serine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (28)...(28)
; OTHER INFORMATION: Xaa = Aspartic Acid or Glutamic Acid

Query Match      74.1%; Score 80; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GHKXNKNKXXAXXKSDTQTQEXXXXXXEE 32
||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1 GHKXNKNKXXAXXKSDTQTQEXXXXXXEE 32

RESULT 12
US-09-853-079-39
; Sequence 39, Application US/09853079
; Publication No. US20030109689A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C11
; CURRENT APPLICATION NUMBER: US/09/853,079
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Babesia microti
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3)...(3)
; OTHER INFORMATION: Xaa = Glycine or Aspartic Acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (5)...(5)
; OTHER INFORMATION: Xaa = Proline or Isoleucine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (7)...(7)
; OTHER INFORMATION: Xaa = Lysine or Threonine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (11)...(11)
; OTHER INFORMATION: Xaa = Glutamic Acid or Glycine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (12)...(12)
; OTHER INFORMATION: Xaa = Lysine or Asparagine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (14)...(14)
; OTHER INFORMATION: Xaa = Glutamic Acid or Glycine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (15)...(15)
; OTHER INFORMATION: Xaa = Isoleucine or Arginine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (18)...(18)
; OTHER INFORMATION: Xaa = Cysteine or Serine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (23)...(23)
; OTHER INFORMATION: Xaa = Threonine or Proline
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (26)...(26)
; OTHER INFORMATION: Xaa = Isoleucine or Threonine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (27)...(27)
; OTHER INFORMATION: Xaa = Cysteine or Serine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (28)...(28)
; OTHER INFORMATION: Xaa = Aspartic Acid or Glutamic Acid
```

```
; OTHER INFORMATION: Xaa = Histidine or Tyrosine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (23)...(23)
; OTHER INFORMATION: Xaa = Threonine or Proline
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (26)...(26)
; OTHER INFORMATION: Xaa = Isoleucine or Threonine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (27)...(27)
; OTHER INFORMATION: Xaa = Cysteine or Serine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (28)...(28)
; OTHER INFORMATION: Xaa = Aspartic Acid or Glutamic Acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (29)...(29)
; OTHER INFORMATION: Xaa = Glutamic Acid or Alanine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (30)...(30)
; OTHER INFORMATION: Xaa = Cysteine or Histidine
US-09-853-079-39

Query Match          74.1%; Score 80; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHXXNKNKXXXXXKSDTQTQXQEXXXXXXEE 32
    |||||
DB 1 GHXXNKNKXXXXXKSDTQTQXQEXXXXXXEE 32

RESULT 13
US-09-737-178-39
; Sequence 39, Application US/09737178
; Patent No. US20010029295A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Babesia microti
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3)...(3)
; OTHER INFORMATION: Xaa = Glycine or Aspartic Acid
; NAME/KEY: VARIANT
; LOCATION: (5)...(5)
; OTHER INFORMATION: Xaa = Proline or Isoleucine
; NAME/KEY: VARIANT
; LOCATION: (7)...(7)
; OTHER INFORMATION: Xaa = Lysine or Threonine
; NAME/KEY: VARIANT
; LOCATION: (11)...(11)
; OTHER INFORMATION: Xaa = Glutamic Acid or Glycine
; NAME/KEY: VARIANT
```

```
; LOCATION: (12)...(12)
; OTHER INFORMATION: Xaa = Lysine or Asparagine
; NAME/KEY: VARIANT
; LOCATION: (14)...(14)
; OTHER INFORMATION: Xaa = Glutamic Acid or Glycine
; NAME/KEY: VARIANT
; LOCATION: (15)...(15)
; OTHER INFORMATION: Xaa = Isoleucine or Arginine
; NAME/KEY: VARIANT
; LOCATION: (18)...(18)
; OTHER INFORMATION: Xaa = Histidine or Tyrosine
; NAME/KEY: VARIANT
; LOCATION: (23)...(23)
; OTHER INFORMATION: Xaa = Threonine or Proline
; NAME/KEY: VARIANT
; LOCATION: (26)...(26)
; OTHER INFORMATION: Xaa = Isoleucine or Threonine
; NAME/KEY: VARIANT
; LOCATION: (27)...(27)
; OTHER INFORMATION: Xaa = Cysteine or Serine
; NAME/KEY: VARIANT
; LOCATION: (28)...(28)
; OTHER INFORMATION: Xaa = Aspartic Acid or Glutamic Acid
; NAME/KEY: VARIANT
; LOCATION: (29)...(29)
; OTHER INFORMATION: Xaa = Glutamic Acid or Alanine
; NAME/KEY: VARIANT
; LOCATION: (30)...(30)
; OTHER INFORMATION: Xaa = Cysteine or Histidine
US-09-737-178-39

Query Match          74.1%; Score 80; DB 10; Length 32;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHXXNKNKXXXXXKSDTQTQXQEXXXXXXEE 32
    |||||
DB 1 GHXXNKNKXXXXXKSDTQTQXQEXXXXXXEE 32

RESULT 14
US-09-286-488-53
; Sequence 53, Application US/09286488
; Patent No. US20020169136A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C3
; CURRENT APPLICATION NUMBER: US/09/286,488
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Babesia microti
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3)...(3)
; OTHER INFORMATION: Xaa = Glycine or Aspartic Acid
; NAME/KEY: VARIANT
; LOCATION: (5)...(5)
; OTHER INFORMATION: Xaa = Proline or Isoleucine
; NAME/KEY: VARIANT
; LOCATION: (7)...(7)
; OTHER INFORMATION: Xaa = Lysine or Threonine
; NAME/KEY: VARIANT
; LOCATION: (11)...(11)
; OTHER INFORMATION: Xaa = Glutamic Acid or Glycine
; NAME/KEY: VARIANT
```

```

US-09-853-079-53
; Sequence 53, Application US/09853079
; Publication NO. US20030109689A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secretst, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C11
; CURRENT APPLICATION NUMBER: US/09/853,079
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-853-079-53

```

```
Query Match      73.1%; Score 79; DB 9; Length 275;
Best Local Similarity 56.2%; Pred. No. 1.7e-08;
Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
```

**Qy** 1 GHXXNXNKSXXXAXKSDTQTQXXXXXXXXEE 32  
||| ||| | ||| ||| |||  
**Db** 216 GHGPNXNKSEKAERKSHDTQTTOEICEECEEE .247

Search completed: July 16, 2003, 18:08:36  
Job time : 6.2344 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:50:14 ; Search time 3.7774 Seconds  
(without alignments)  
814.396 Million cell updates/sec

Title: US-09-853-079-39

Perfect score: 108

Sequence: 1 GHKXNKNKSSXXAXKSDTQTQXQXXXXXEE 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	36.1	1431	2 A15866	dextranucrase (EC
2	38.5	35.6	1033	2 S19247	cell adhesion prot
3	38	35.2	488	2 B69825	cell wall-binding
4	37	34.3	290	2 A82236	pseudouridine synt
5	37	34.3	367	1 F64232	hypothetical prote
6	36	33.3	251	2 S59314	hypothetical prote
7	35	32.4	261	2 C84584	probable CAMP-depe
8	35	32.4	425	2 C86232	hypothetical prote
9	35	32.4	1227	2 T23004	hypothetical prote
10	34	31.5	82	2 T45054	hypothetical prote
11	34	31.5	292	2 T28817	hypothetical prote
12	34	31.5	368	2 C72260	hypothetical prote
13	34	31.5	492	2 T01086	probable serine/th
14	34	31.5	597	2 D64711	glutamine-fructose
15	34	31.5	597	2 H71809	glutamine-fructose
16	34	31.5	618	2 T21507	hypothetical prote
17	34	31.5	960	2 S54461	hypothetical prote
18	34	31.5	1048	2 S64758	hypothetical prote
19	34	31.5	1141	2 T29104	SCD25 protein (ver
20	34	31.5	1166	2 C96945	Tbcl protein - mou
21	34	31.5	1250	2 S14177	cation efflux syst
22	34	31.5	1400	2 T52359	SCD25 protein (ver
23	34	31.5	1546	2 G90603	lipoprotein (impor
24	33	30.6	127	1 KNMUHY	dehydrin-like prot
25	33	30.6	127	2 T45729	dehydrin-like prot
26	33	30.6	201	2 S16286	opacity protein op
27	33	30.6	327	2 T51571	hypothetical prote
28	33	30.6	370	2 A89782	conserved hypothet
29	33	30.6	431	2 A37801	IstA protein homol

30 33 30.6 483 2 A69745  
31 33 30.6 543 1 QFMSL  
32 33 30.6 562 2 S69055  
33 33 30.6 571 2 H84798  
34 33 30.6 605 2 S46833  
35 33 30.6 805 2 A56199  
36 33 30.6 1966 2 T08991  
37 32 29.6 91 2 G90064  
38 32 29.6 117 2 S71595  
39 32 29.6 139 2 G71362  
40 32 29.6 165 2 T14462  
41 32 29.6 183 2 T43439  
42 32 29.6 217 2 S10212  
43 32 29.6 253 2 F95955  
44 32 29.6 266 1 A35037  
45 32 29.6 272 2 C44816

hypothetical prote  
neurofilament trip  
TTAGGG repeat-bind  
hypothetical prote  
hypothetical prote  
transcription fact  
hypothetical prote  
hypothetical prote  
triose-phosphate i  
hypothetical prote  
non intermediate f  
hypothetical prote  
late 33K protein -  
hypothetical expor  
insulin-like growt  
N-acetylmuramoyl-L

#### ALIGNMENTS

##### RESULT 1

A45866

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans

C:Species: Streptococcus mutans

C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 07-May-1999

C:Accession: A45866

R:Honda, O.; Kato, C.; Kuramitsu, H. K.

J. Gen. Microbiol. 136, 2099-2105, 1990

A:Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyltransferase

A:Reference number: A45866; MUID:91100958; PMID:2148600

A:Accession: A45866

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1431 <HON>

A:Cross-references: GB:M29296

C:Superfamily: cpl repeat homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:181-201/Domain: cpl repeat homology <CP1>

F:1127-1146/Domain: cpl repeat homology <CP2>

F:1192-1211/Domain: cpl repeat homology <CP3>

F:1257-1276/Domain: cpl repeat homology <CP4>

F:1277-1297/Domain: cpl repeat homology <CP5>

F:1321-1340/Domain: cpl repeat homology <CP8>

F:1341-1361/Domain: cpl repeat homology <CP6>

F:1385-1404/Domain: cpl repeat homology <CP7>

Query Match 36.1%; Score 39; DB 2; Length 1431;

Best Local Similarity 27.6%; Pred. No. 10;

Matches 8; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 4 KXNKNKSSXXAXKSDTQTQXQXXXXXEE 32

DB 106 QANSDDKTVTNTKSEEAQTSEERTKQSEE 134

##### RESULT 2

S19247

cell adhesion protein Gpl60-Dtrk - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 24-Sep-1999

C:Accession: S19247

R:Pulido, D.; Campuzano, S.; Koda, T.; Modolell, J.; Barbacid, M.

EMBO J. 11, 391-404, 1992

A:Title: Dtrk, a Drosophila gene related to the trk family of neurotrophin receptors,

A:Reference number: S19247; MUID:92164624; PMID:1371458

A:Accession: S19247

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1033 <PUL>

A:Cross-references: EMBL:X63453; NID:g7883; PIDN:CAA45053.1; PID:g7884

C:Genetics:

A:Gene: FlyBase:TK48D

A:Cross-references: FlyBase:FBgn0004839  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  
C:Keywords: ATP  
F:690-1028/Domain: protein kinase homology <KIN>  
F:698-706/Region: protein kinase ATP-binding motif

Query Match 35.6%; Score 38.5; DB 2; Length 1033;  
Best Local Similarity 40.0%; Pred. No. 9.4;  
Matches 10; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

QY 1 GHXXNXXNKKXXAXXKSDTQTQX 25  
||| : : : : :  
DB 648. GHSK-SRSGSGDAQKSDDTACSQQ 671

RESULT 3  
B69825  
cell wall-binding protein homolog yhdD - Bacillus subtilis  
C:Species: Bacillus subtilis  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C:Accession: B69825  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berta  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
teich, J.; Harwood, C.R.; Hentaut, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, F.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, M  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: B69825  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-488 <KUN>  
A:Cross-references: GB:AL009126; NID:g2633260; PIDN:CAB12776.1; PID:ell182938;  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yhdB

Query Match 35.2%; Score 38; DB 2; Length 488;  
Best Local Similarity 36.8%; Pred. No. 5.6;  
Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 4 KXXNXXKXXAXXKSDTQTQ 22  
||| : : : : :  
DB 148 KSNKSSSSSSKSSSSNKS 166

RESULT 4  
A82236  
pseudouridine synthase family 1 protein VC1140 [imported] - Vibrio cholerae (strain N169  
C:Species: Vibrio cholerae  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: A82236  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H  
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: A82236  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-290 <HEI>  
A:Cross-references: GB:AE004194; GB:AE003852; NID:g9655612; PIDN:AAF94299.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:  
A:Gene: VC1140  
A:Map position: 1

Query Match 34.3%; Score 37; DB 2; Length 290;  
Best Local Similarity 36.4%; Pred. No. 5.4;  
Matches 8; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

QY 1 GHXXNXXNKKXXAXXKSDTQT 22  
||| : : : : :  
DB 79 GHPTPRANRKSXVANKKNATQT 100

RESULT 5  
F64232  
hypothetical protein homolog MG295 - Mycoplasma genitalium  
C:Species: Mycoplasma genitalium  
C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 07-Dec-1999  
C:Accession: F64232  
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.  
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,  
C.A.; Venter, J.C.  
Science 270, 397-403, 1995  
A:Title: The minimal gene complement of Mycoplasma genitalium.  
A:Reference number: A64200; MUID:96026346; PMID:7569993  
A:Accession: F64232  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-367 <TIGR>  
A:Cross-references: GB:U39710; GB:LA43967; NID:gl045989; PID:gl045993; TIGR:MG295  
A:Experimental source: strain G-37  
C:Genetics:  
A:Genetic code: SGC3  
C:Superfamily: probable membrane protein YDL033C

Query Match 34.3%; Score 37; DB 1; Length 367;  
Best Local Similarity 70.0%; Pred. No. 6.9;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GHXXNXXNKS 10  
||| : : : : :  
DB 50 GHKKNNKS 59

RESULT 6  
S59314  
hypothetical protein YLR126c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein L3105  
C:Species: Saccharomyces cerevisiae  
C>Date: 29-Nov-1995 #sequence\_revision 23-Feb-1996 #text\_change 19-Apr-2002  
C:Accession: S59314; S64963; S64968; S69413  
R:Delius, H.  
submitted to the EMBL Data Library, June 1995  
A:Description: 36.8 kb of S.cerevisiae chromosome XII including ACE2, CK11, PDC5, SL5  
A:Reference number: S59313  
A:Accession: S59314  
A:Molecule type: DNA  
A:Residues: 1-251 <DEL>  
A:Cross-references: EMBL:X01258; NID:g995686; PIDN:CAA62637.1; PID:g995688  
A:Experimental source: strain S288C  
R:Verhasselt, P.; Voet, M.; Voickaert, G.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64943  
A:Accession: S64963  
A:Molecule type: DNA  
A:Residues: 1-251 <VER>  
A:Cross-references: EMBL:Z73298; NID:gl360532; PIDN:CAA97695.1; PID:e245566; PID:gl36  
A:Experimental source: strain S288C  
R:Delius, H.; Hebling, U.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64967  
A:Accession: S64968  
A:Molecule type: DNA

A:Residues: 1-251 <DEW>

A:Cross-references: EMBL:Z73298; NID:gl1360532; PIDN:CAA97695.1; PID:G245566; PID:gl136053

A:Experimental source: strain S288C

R:Verhasselt, P.; Volckaert, G.

submitted to the EMBL Data Library, September 1995

A:Reference number: S69393

A:Accession: S69413

A:Molecule type: DNA

A:Residues: 1-251 <VEW>

A:Cross-references: EMBL:X89514; NID:gl1297019; PIDN:CAA61704.1; PID:el98755; PID:gl129704

C:Genetics:

A:Cross-references: SGD:S0004116

A:Map position: 128

Query Match 33.3%; Score 36; DB 2; Length 251;

Best Local Similarity 32.0%; Pred. No. 7.6;

Matches 8; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

QY 1 GHXXNXXKXXKXXKXDTQTXQE 25

|| : : : || : : : || : : : || : : : || : : : ||

DB 197 GHPEFNSDVAQKGLKLSQDKLTLEE 221

RESULT 7

C84584

probable cAMP-dependent protein kinase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C:Accession: C84584

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: C84584

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-261 <STO>

A:Cross-references: GB:AE002093; NID:g4580468; PIDN:AAD24392.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g20040

A:Map position: 2

Query Match 32.4%; Score 35; DB 2; Length 261;

Best Local Similarity 32.3%; Pred. No. 13;

Matches 10; Conservative 1; Mismatches 20; Indels 0; Gaps 0;

QY 2 HXXNXXKXXKXXKXDTQTXQEXXXXXE 32

|| : : : || : : : || : : : || : : : || : : : ||

DB 230 HPLENDNVLPLETSKSLDTEDQDAQNWLEE 260

RESULT 8

C86232

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: C86232

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: C86232

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-425 <STO>

A:Cross-references: GB:AE005172; NID:g2160172; PIDN:AABG0735.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 32.4%; Score 35; DB 2; Length 425;

Best Local Similarity 30.0%; Pred. No. 21;

Matches 6; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 6 NXNXXKXXKXXKXDTQTXQE 25

|| : : : || : : : || : : : || : : : || : : : ||

DB 197 NGNKKLDAFRDAETKTLED 216

RESULT 9

T23004

hypothetical protein F59F3.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T23004

R:Kershaw, J.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z19651

A:Accession: T23004

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1227 <WIL>

A:Cross-references: EMBL:Z68005; PIDN:CAA91990.1; GSPDB:GN00028; CESP:F59F3.1

C:Experimental source: clone F59F3

C:Genetics:

A:Gene: CESP:F59F3.1

A:Map position: X

A:Introns: 31/3; 58/1; 279/3; 429/2; 544/3; 638/2; 801/2; 887/1; 929/2; 1025/3; 1062/

Query Match

Best Local Similarity 40.0%; Pred. No. 60;

Matches 8; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 1 GHXXNXXKXXKXXKXSDT 20

|| : : : || : : : || : : : || : : : || : : : ||

DB 1050 GLSKKHTNKTYRTKSKDT 1069

RESULT 10

T45054

hypothetical protein Y39B6B.bb [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000

C:Accession: T45054

R:Wilson, R.; Alnscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto

raser, A.; Fulton, L.; Gardner, A.; Green, P.; Hawkins, T.; Hillier, L.; Jier, M.; Jo

B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Rifken, L.; Roopra, A.; Saunders, D.

Nature 388, 32-38, 1994

A:Authors: Shownkeen, R.; Sims, M.; Smaldon, N.; Smith, A.; Smith, M.; Sonhammer, E.

tock, L.; Wilkinson-Sproat, J.; Wohldman, P.

A:Title: 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.

A:Reference number: S43531; MUID:94150718; PMID:7906398

A:Accession: T45054

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-82 <WIL>

A:Cross-references: EMBL:AL132896; PIDN:CAB60933.1

A:Experimental source: clone Y39B6B

C:Genetics:

A:Map position: 3

A:Introns: 26/2; 67/1

A:Note: Y39B6B.bb

Query Match

Best Local Similarity 37.5%; Score 34; DB 2; Length 82;

Matches 9; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 GHXXNXXKXXKXXKXSDTQTXQ 24

Db 53 GHRVQNKSGQATDKSDQFQQQ 76

RESULT 11

T28817

hypothetical protein F07C3.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 04-Mar-2000

C:Accession: T28817

R:Favella, A.; Gattung, S.

submitted to the EMBL Data Library, March 1996

A:Description: The sequence of C. elegans cosmid F07C3.

A:Reference number: Z20528

A:Accession: T28817

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-292 <FAV>

A:Cross-references: EMBL:U50308; PIDN:AAC48005.1; GSPDB:GN00023; CESP:F07C3.5

A:Experimental source: strain Bristol N2; clone F07C3

C:Genetics:

A:Gene: CESP:F07C3.5

A:Map position: 5

A:Introns: 54/1; 136/1; 160/3; 210/3; 283/3

C:Superfamily: Caenorhabditis elegans hypothetical protein F07C3.5

Query Match 31.5%; Score 34; DB 2; Length 292;

Best Local Similarity 31.8%; Pred. No. 23;

Matches 7; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 4 KXNKNKXXXAXXKSDTQTQXOE 25

Db 4 KRSTNSRKTANSSSDTSTSE 25

RESULT 12

C72260

hypothetical protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: C72260

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: C72260

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-368 <ARN>

A:Cross-references: GB:AE001791; GB:AE000512; NID:g4981929; PIDN:AAD36441.1; PID:g498193

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1370

Query Match 31.5%; Score 34; DB 2; Length 368;

Best Local Similarity 47.1%; Pred. No. 29;

Matches 8; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 GHXXKNKSKXXXKXS 17

Db 207 GHILVLKLNKSRADMK 223

RESULT 13

T01086

probable serine/threonine-specific protein kinase (EC 2.7.1.-) T10P11.10 - Arabidopsis

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 24-Mar-1999

C:Accession: T01086

R:Kaplan, N.; Johnson, D.; Schutz, K.; Gnoj, L.; Hoffman, J.; Till, S.; de la Bastide, M

hi, M.; Martienssen, R.; Chen, E.Y.; Wilson, R.; McCombie, W.R.

submitted to the EMBL Data Library, November 1998

A:Description: Sequence of A. thaliana BAC T10P11 from chromosome IV.

A:Reference number: Z14248

A:Accession: T01086

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-492 <RAP>

A:Cross-references: EMBL:AC002330; NID:g2262135; PID:g2262143

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Note: T10P11.10

C:Keywords: phosphotransferase

Query Match 31.5%; Score 34; DB 2; Length 492;

Best Local Similarity 36.8%; Pred. No. 39;

Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 2 HXKNXNKSXXXKSKXDT 20

Db 47 YRRSNKNKSLESSKSNHT 65

RESULT 14

D64711

glutamine-fructose-6-phosphate transaminase (isomerizing) (EC 2.6.1.16) - Helicobacte

N:Alternate names: glucosamine fructose-6-phosphate aminotransferase [misnomer]

C:Species: Helicobacter pylori

C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 19-Jul-2002

C:Accession: D64711

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser,

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: D64711

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-597 <TOM>

A:Cross-references: GB:AE000651; GB:AE000511; NID:g2314708; PIDN:AAD08570.1; PID:g231

C:Superfamily: glutamine-fructose-6-phosphate aminotransferase (isomerizing)

C:Keywords: aminotransferase; intramolecular oxidoreductase; isomerase

F;2-597/Product: glutamine-fructose-6-phosphate transaminase (isomerizing) #status pr

F;2/Active site: Cys #status predicted

Query Match 31.5%; Score 34; DB 2; Length 597;

Best Local Similarity 33.3%; Pred. No. 47;

Matches 8; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 2 HXKNXNKSXXXKSKXDTQTQXOE 25

Db 326 HFKNPNELFIALSQSGTADTLE 349

RESULT 15

H71809

glutamine-fructose-6-phosphate transaminase (isomerizing) (EC 2.6.1.16) - Helicobacte

C:Species: Helicobacter pylori

A:Variety: Strain J99

C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 19-Jul-2002

C:Accession: H71809

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: H71809

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-597 <ARN>

A:Cross-references: GB:AE001564; GB:AE001439; NID:g4156032; PIDN:AAD06999.1; PID:g415



A:Experimental source: strain J99  
 C:Genetics:  
 A:Gene: glms  
 C:Superfamily: glutamine-fructose-6-phosphate aminotransferase (isomerizing)  
 C:Keywords: aminotransferase; intramolecular oxidoreductase; isomerase  
 F:2-597/Product: glutamine-fructose-6-phosphate transaminase (isomerizing) #status predi  
 F:2/Active site: Cys #status predicted

Query Match 31.5%; Score 34; DB 2; Length 597;  
 Best Local Similarity 33.3%; Pred. No. 47;  
 Matches 8; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 2 HXXNXXKXXAXXKXDTQTXQE 25  
 | | | | : | : | |  
 Db 326 HFKSNPNELFTAISQGETADTLE 349

Search completed: July 16, 2003, 18:09:56  
 Job time : 5.7774 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:37:24 ; Search time 1.8887 Seconds  
(without alignments)  
702.728 Million cell updates/sec

Title: US-09-853-079-39

Perfect score: 108

Sequence: 1 GHXXNXXNXXKXXKSDTQTQXQXXXXXEE 32

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	39	36.1	1462	1	GTFD_STRMU
2	37	34.3	367	1	TRMU_MYCGE
3	36	33.3	503	1	PODX_MOUSE
4	34	31.5	119	1	TAT_SIVAI
5	34	31.5	596	1	GLMS_HELPJ
6	34	31.5	596	1	GLMS_HELPJ
7	34	31.5	960	1	YMX6_YEAST
8	34	31.5	1253	1	SC25_YEAST
9	33	30.6	128	1	DX11_ARATH
10	33	30.6	431	1	T232_BACTB
11	33	30.6	542	1	NFL_MOUSE
12	33	30.6	562	1	TBF1_YEAST
13	33	30.6	605	1	APM2_YEAST
14	33	30.6	805	1	E2F_DROME
15	32	29.6	217	1	V33P_ADE41
16	32	29.6	266	1	IBP3_PIG
17	32	29.6	272	1	CWLK_BACSU
18	32	29.6	277	1	YKC3_YEAST
19	32	29.6	291	1	IBP3_BOVIN
20	32	29.6	231	1	IBP3_HUMAN
21	32	29.6	323	1	YCJ5_SCHPO
22	32	29.6	807	1	OSB1_HUMAN
23	32	29.6	809	1	OSB1_RABIT
24	32	29.6	815	1	YC53_YEAST
25	32	29.6	946	1	YIN7_YEAST
26	32	29.6	1026	1	MTLB_DROME
27	32	29.6	1049	1	NMD2_SCHPO
28	32	29.6	1241	1	TRK1_SACBA
29	32	29.6	3178	1	YS89_CAEEL
30	31.5	29.2	651	1	PBP2_HAEIN
31	31	28.7	160	1	SSB_CHLMU
32	31	28.7	198	1	ROVA_STRPY
33	31	28.7	205	1	RS4_RICPR

#### RESULT 1

ID	GTFD_STRMU	STANDARD;	PRT;	1462 AA.
AC	P49331; O69383; O69386; O69389; O69392; O69398;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)			
DE	(Sucrose 6-glucosyltransferase).			
GN	GTFD.			
OS	Streptococcus mutans.			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1309;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GS-5;			
RX	MEDLINE=91100958; PubMed=2148600;			
RA	Honda O., Kato C., Kuramitsu H.K.;			
RT	"Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyltransferase-S enzyme."			
RL	J. Gen. Microbiol. 136:2099-2105(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MT4239, MT4245, MT4251, MT4467, and MT8148;			
RX	MEDLINE=98231643; PubMed=9570124;			
RA	Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S., Kimura S., Hamada S.;			
RT	"Molecular analyses of glucosyltransferase genes among strains of Streptococcus mutans."			
RL	FEMS Microbiol. Lett. 161:331-336(1998).			
CC	-1- FUNCTION: PRODUCTION OF EXTRACELLULAR GLUCANS, THAT ARE THOUGHT TO PLAY A KEY ROLE IN THE DEVELOPMENT OF THE DENTAL PLAQUE BECAUSE OF THEIR ABILITY TO ADHERE TO SMOOTH SURFACES AND MEDIATE THE AGGREGATION OF BACTERIAL CELLS AND FOOD DEBRIS.			
CC	-1- CATALYTIC ACTIVITY: Sucrose + [(1,6)-alpha-D-glucosyl](N) - D-fructose + [(1,6)-alpha-D-glucosyl](N+1).			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- DISEASE: DENTAL CARIES.			
CC	-1- MISCELLANEOUS: GTF-1 SYNTHESIZES WATER-INSOLUBLE GLUCANS (ALPHA 1,3-LINKED GLUCOSE AND SOME 1,6 LINKAGES), GTF-S SYNTHESIZES WATER-SOLUBLE GLUCANS (ALPHA 1,6-GLUCOSE). GTF-SI SYNTHESIZES BOTH FORMS OF GLUCANS.			
CC	-1- SIMILARITY: TO OTHER GLUCOSYLTRANSFERASES AND SOME TO A GLUCAN-BINDING PROTEIN FROM S.MUTANS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M29296; AAA26895.1; -			
DR	EMBL; D88653; BAA26103.1; -			

```

DR EMBL; D88656; BAA26107.1; -
DR EMBL; D88659; BAA26111.1; -
DR EMBL; D88662; BAA26115.1; -
DR EMBL; D89979; BAA26121.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; GH_70.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase; Glycosyltransferase; Signal; Repeat; Dental caries.
FT SIGNAL ?
FT CHAIN ?
FT DOMAIN 1232 1423
FT REPEAT 1232 1295
FT REPEAT 1296 1359
FT REPEAT 1360 1423
FT VARIANT 58 58
FT VARIANT 68 68
FT VARIANT 81 81
FT VARIANT 113 113
FT VARIANT 122 122
FT VARIANT 132 132
FT VARIANT 135 135
FT VARIANT 202 202
FT VARIANT 255 255
FT VARIANT 275 275
FT VARIANT 288 288
FT VARIANT 301 301
FT VARIANT 313 313
FT VARIANT 317 317
FT VARIANT 328 328
FT VARIANT 350 350
FT VARIANT 628 633
FT VARIANT 688 688
FT VARIANT 726 732
FT VARIANT 726 730
FT VARIANT 964 964
FT VARIANT 1019 1019
FT VARIANT 1059 1060
FT VARIANT 1060 1060
FT VARIANT 1080 1080
FT VARIANT 1142 1142
FT VARIANT 1198 1198
FT VARIANT 1220 1220
FT VARIANT 1280 1280
FT VARIANT 1282 1282
FT VARIANT 1290 1290
FT VARIANT 1311 1311
FT VARIANT 1403 1403
FT VARIANT 1425 1425
FT VARIANT 1449 1449
FT CONFLICT 1428 1462
SQ SEQUENCE 1462 AA; 163512 MW; 5C6541F0DCB0DF00 CRC64;

Query Match 36.1%; Score 39; DB 1; Length 1462;
Best Local Similarity 27.6%; Pred. No. 4.2;
Matches 8; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 4 KXNKNXXKXKXKXDTQTOEXXXXXXEE 32
Db 106 QANSDDKTVTNKEEAQTSEERTKQSEE 134

RESULT 2

```

```

TRMU_MYCGE
ID TRMU_MYCGE STANDARD; PRT; 367 AA.
AC P47537;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase
DE (EC 2.1.1.61).
GN TRMU OR MG295.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RC SEQUENCE FROM N.A.
RA STRAIN=ATCC 33530 / G-37;
RA MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA "the minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing 5-methylaminomethyl-2-
CC thioridylate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE TRMU FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39709; AAC71516.1; -
CC TIGR; MG295; -
CC InterPro; IPR004506; Trmu.
CC InterPro; IPR004135; tRNA_Me_trans.
CC Pfam; PF03054; tRNA_Me_trans; 1.
CC TIGRFAMs; TIGR00420; trmu; 1.
CC Transferase; Methyltransferase; tRNA processing; Complete proteome.
SQ SEQUENCE 367 AA; 41887 MW; EBC6409D8C2D0625 CRC64;

Query Match 34.3%; Score 37; DB 1; Length 367;
Best Local Similarity 70.0%; Pred. No. 2.6;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GHKXNKNKS 10
Db 50 GHKXNKNKS 59

RESULT 3
PODX_MOUSE
ID PODX_MOUSE STANDARD; PRT; 503 AA.
AC Q9K0M4; Q9ESZ1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Podocalyxin-like protein 1 precursor.
GN PODXL OR PCPL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RA Hara T., Nakano Y., Tanaka M., Tamura K., Sekiguchi T., Minehata K.,

```

RA Copeland N.G., Jenkins N.A., Okabe M., Kogo H., Mukoyama Y.,  
RA Miyajima A.;  
RT "Identification of podocalyxin-like protein 1 as a novel cell surface  
RT marker for hemangioblasts in the murine aorta-gonad-mesonephros  
RT region";  
RL Immunity 11:567-578(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kershaw D.B., Li J.;  
RT "Gene structure of mouse podocalyxin";  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: Functions as an antiadhesin that maintains an open  
CC filtration pathway between neighboring foot processes in the  
CC podocyte by charge repulsion.  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -!- TISSUE SPECIFICITY: Glomerular epithelium cell (podocyte).  
CC -!- PTM: Sialoglycoprotein (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE PODOCALYXIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB028048; BAA86912.1; -;  
DR EMBL; AF290209; AAG02458.1; -;  
DR MGD; MGI:1351317; Podxl.  
KW Glycoprotein; Signal; Transmembrane; Polymorphism.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 503 PODOCALYXIN-LIKE PROTEIN 1.  
FT DOMAIN 22 404 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 405 425 POTENTIAL.  
FT DOMAIN 426 503 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 27 259 SER/THR-RICH.  
FT CARBOHYD 25 25 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 167 167 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 78 78 S -> F.  
SQ SEQUENCE 503 AA; 53389 MW; 786ALECF65484DIF CRC64;  
  
Query Match 33.3%; Score 36; DB 1; Length 503;  
Best Local Similarity 26.9%; Pred. No. 5.9;  
Matches 7; Conservative 4; Mismatches 15; Indels 0; Gaps 0;  
  
QY 6 NXNKSXXAXKSDTQTQEXXXXXX 31  
| | : : | | : |  
Db 23 NGNETSTSAIKSTVSHQSATTSTE 48  
  
RESULT 4  
TAT\_SIVAL  
ID TAT\_SIVAL STANDARD; PRT; 119 AA.  
AC P27975;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE TAT protein (Transactivating regulatory protein).  
GN TAT.  
OS Simian immunodeficiency virus (AGM155 isolate) (SIV-AGM).  
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90156504; PubMed=2304139;  
RA Johnson P.R., Fomsgaard A., Allan J., Gravell M., London W.T.,

RA Olmstead R.A., Hirsch V.M.;  
RT "Simian immunodeficiency viruses from African green monkeys display  
RT unusual genetic diversity";  
RL J. Virol. 64:1086-1092(1990).  
CC -!- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS BY BINDING TO THE  
CC TRANS-ACTIVATING RESPONSE SEQUENCE (TAR) RNA ELEMENT AND  
CC ACTIVATES TRANSCRIPTION INITIATION AND/OR ELONGATION FROM THE LTR  
CC PROMOTER.  
CC -!- SUBUNIT: BINDS CYCLIN T1 (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC -!- MISCELLANEOUS: THE 155 ISOLATE IS FROM A MONKEY IMPORTED FROM  
CC KENYA.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M29975; AAA91909.1; -;  
DR EMBL; IPR001831; HIV\_Tat.  
DR Pfam; PF00539; Tat; 1.  
KW Transcription regulation; Activator; RNA-binding; Nuclear protein;  
KW AIDS.  
SQ SEQUENCE 119 AA; 13648 MW; 958DB1A3F7027729 CRC64;  
  
Query Match 31.5%; Score 34; DB 1; Length 119;  
Best Local Similarity 41.2%; Pred. No. 3.5;  
Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;  
  
QY 16 KSXDTQTQEXXXXXX 32  
| | : | | | | |  
Db 78 RGRDSQTQESQKVEE 94  
  
RESULT 5  
GLMS\_HELPJ  
ID GLMS\_HELPJ STANDARD; PRT; 596 AA.  
AC Q9ZJ94;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Glucosamine--fructose-6-phosphate aminotransferase [isomerizing-  
DE (EC 2.6.1.16) (Hexosephosphate aminotransferase) (D-fructose-6-  
DE phosphate amidotransferase) (GFAT) (L-glutamine-D-fructose-6-phosphate  
DE amidotransferase) (Glucosamine-6-phosphate synthase).  
DE GLMS OR JHP1420.  
GN Helicobacter pylori J99 (campylobacter pylori J99).  
OS Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OX NCBI\_TaxID=85963;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99120557; PubMed=9923682;  
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,  
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
RA Trust T.J.;  
RT "Genomic sequence comparison of two unrelated isolates of the human  
RT gastric pathogen Helicobacter pylori";  
RL Nature 397:176-180(1999)  
CC -!- FUNCTION: CATALYZES THE FIRST STEP IN HEXOSAMINE METABOLISM, A  
CC CONVERTING FRUCTOSE-6P INTO GLUCOSAMINE-6P USING GLUTAMINE AS A  
CC NITROGEN SOURCE (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: L-glutamine + D-fructose 6-phosphate = L-  
CC glutamate + D-glucosamine 6-phosphate.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.  
CC GFAT SUBFAMILY.  
CC -!- SIMILARITY: CONTAINS 1 TYPE-2 GLUTAMINE AMIDOTRANSFERASE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE001564; AAD06999.1; -  
 DR HSSP; P17169; IGDO.  
 DR InterPro; IPR000583; GATase\_2.  
 DR InterPro; IPR001347; SIS.  
 DR Pfam; PF00310; GATase\_2; 1.  
 DR Pfam; PF01380; SIS; 2.  
 DR TIGRFAMs; TIGR01135; glms; 1.  
 DR PROSITE; PS00443; GATASE\_TYPE\_II; 1.  
 KW Transferase; Amino transferase; Glutamine amidotransferase;  
 KW Complete proteome.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 591 591 ISOMERIZATION FRU-6P (BY SIMILARITY).  
 FT DOMAIN 1 181 GLUTAMINE AMIDOTRANSFERASE.  
 FT SEQUENCE 596 AA; 66926 MW; C438BCCB6A245C28 CRC64;

Query Match 31.5%; Score 34; DB 1; Length 596;  
 Best Local Similarity 33.3%; Pred. No. 18;  
 Matches 8; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 2 HXKXNKNKXXAXXKSDTQTQX 25  
 DB 325 HFKSNPNELFAISQSGETADTLE 348

## RESULT 6

GLMS\_HELPY  
 ID GLMS\_HELPY STANDARD; PRT; 596 AA.  
 AC 026060;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]  
 DE (EC 2.6.1.16) (Hexosephosphate aminotransferase) (D-fructose-6-  
 DE phosphate amidotransferase) (GFAT) (L-glutamine-D-fructose-6-phosphate  
 DE amidotransferase) (Glucosamine-6-phosphate synthase).  
 GN GLMS OR HP1532.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 OC Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.W., Fujii C., Bowman C., Watthey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen Helicobacter  
 RT pylori."  
 RL Nature 388:539-547(1997).  
 CC -!- FUNCTION: CATALYZES THE FIRST STEP IN HEXOSAMINE METABOLISM,  
 CC CONVERTING FRUCTOSE-6P INTO GLUCOSAMINE-6P USING GLUTAMINE AS A  
 CC NITROGEN SOURCE (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: L-glutamine + D-fructose 6-phosphate = L-  
 CC glutamate + D-glucosamine 6-phosphate.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.

CC GFAT SUBFAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 TYPE-2 GLUTAMINE AMIDOTRANSFERASE DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE000651; AAD08570.1; -  
 DR HSSP; P17169; IGDO.  
 DR MEROPS; C44.971; -  
 DR TIGR; HPI532; -  
 DR InterPro; IPR000583; GATase\_2.  
 DR InterPro; IPR001347; SIS.  
 DR Pfam; PF00310; GATase\_2; 1.  
 DR Pfam; PF01380; SIS; 2.  
 DR TIGRFAMs; TIGR01135; glms; 1.  
 DR PROSITE; PS00443; GATASE\_TYPE\_II; 1.  
 KW Transferase; Amino transferase; Glutamine amidotransferase;  
 KW Complete proteome.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 GATASE (BY SIMILARITY).  
 FT ACT\_SITE 591 591 ISOMERIZATION FRU-6P (BY SIMILARITY).  
 FT DOMAIN 1 181 GLUTAMINE AMIDOTRANSFERASE.  
 FT SEQUENCE 596 AA; 66939 MW; 089C534BB85AD3F1 CRC64;

Query Match 31.5%; Score 34; DB 1; Length 596;  
 Best Local Similarity 33.3%; Pred. No. 18;  
 Matches 8; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 2 HXKXNKNKXXAXXKSDTQTQX 25  
 DB 325 HFKSNPNELFAISQSGETADTLE 348

## RESULT 7

YMX6\_YEAST  
 ID YMX6\_YEAST STANDARD; PRT; 960 AA.  
 AC Q04279;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Hypothetical 105.9 kDa protein in ADH3-RCAL intergenic region.  
 GN YMR086W OR YMR582.10.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RA Gentles S., Bowman S., Barrell B.G., Rajandream M.A.;  
 RA Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: SOME, TO YEAST YKL105C.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z49259; CAA89232.1; -  
 DR SGD; S0004692; YMR086W.  
 KW Hypothetical protein.  
 SQ SEQUENCE 960 AA; 105873 MW; 05A4FA27129DB09B CRC64;  
 Query Match 31.5%; Score 34; DB 1; Length 960;  
 Best Local Similarity 33.3%; Pred. No. 30;







RX MEDLINE=91060592; PubMed=2246261;  
RA Nakahira K., Ikenaka K., Wada K., Tamura T.A., Furuichi T.,  
RT Mikoshiba K.;  
RT "Structure of the 68-kDa neurofilament gene and regulation of its  
expression";  
RL J. Biol. Chem. 265:19786-19791(1990).  
CC -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,  
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.  
CC -!- DOMAIN: THE EXTRA MASS AND HIGH CHARGE DENSITY THAT DISTINGUISH  
CC THE NEUROFILAMENT PROTEINS FROM ALL OTHER INTERMEDIATE FILAMENT  
CC PROTEINS ARE DUE TO THE TAILPIECE EXTENSIONS. THIS REGION MAY FORM  
CC A CHARGED SCAFFOLDING STRUCTURE SUITABLE FOR INTERACTION WITH  
CC OTHER NEURONAL COMPONENTS OR IONS.  
CC -!- MISCELLANEOUS: NF-L IS THE MOST ABUNDANT OF THE THREE  
CC NEUROFILAMENT PROTEINS AND, AS THE OTHER NONEPITHELIAL  
CC INTERMEDIATE FILAMENT PROTEINS, IT CAN FORM HOMOPOLYMERIC 10-NM  
CC FILAMENTS.  
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X02165; CAB51616.1; -  
DR EMBL; M20480; AAA39814.1; -  
DR EMBL; M13016; AAA39810.1; -  
DR EMBL; M55423; AAA39812.1; -  
DR PIR; A25227; QFMSL.  
DR MGD; MGI:97313; NFI.  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; filament; 1.  
DR PROSITE; PS00226; IF; 1.  
KW Intermediate filament; Coiled coil; Neurone; Glycoprotein.  
FT INIT MET 0 0  
FT DOMAIN 1 92 HEAD.  
FT DOMAIN 93 396 TAIL, SUBDOMAIN B (ACIDIC).  
FT DOMAIN 397 542 ROD.  
FT DOMAIN 93 124 COIL 1A.  
FT DOMAIN 125 137 LINKER 1.  
FT DOMAIN 138 233 COIL 1B.  
FT DOMAIN 234 252 LINKER 12.  
FT DOMAIN 253 271 COIL 2A.  
FT DOMAIN 272 280 LINKER 2.  
FT DOMAIN 281 396 COIL 2B.  
FT DOMAIN 397 443 TAIL, SUBDOMAIN A.  
FT DOMAIN 444 542 TAIL, SUBDOMAIN B (ACIDIC).  
FT CARBOHYD 20 20 O-LINKED (GLCNAC) (BY SIMILARITY).  
FT CARBOHYD 26 26 O-LINKED (GLCNAC) (BY SIMILARITY).  
FT SITE 381 391 EPITOPE (RECOGNIZED BY IF-SPECIFIC  
FT MONOCLONAL ANTIBODY).  
FT CONFLICT 5 5 Y -> S (IN REF. 1).  
FT CONFLICT 8 8 Y -> I (IN REF. 1).  
FT CONFLICT 64 64 M -> K (IN REF. 1).  
FT CONFLICT 72 72 V -> L (IN REF. 2).  
FT CONFLICT 98 98 D -> H (IN REF. 1).  
FT CONFLICT 194 194 R -> A (IN REF. 1).  
FT CONFLICT 202 202 MISSING (IN REF. 2).  
FT CONFLICT 239 239 Y -> I (IN REF. 1).  
SQ SEQUENCE 542 AA; 61448 MW; 8EE9B8C6F0831D8C CRC64;  
  
Query Match 30.6%; Score 33; DB 1; Length 542;  
Best Local Similarity 25.0%; Pred. No. 27;  
Matches 8; Conservative 5; Mismatches 19; Indels 0; Gaps 0;  
  
QY 1 GHKXNNKNSXXAXKSDTQTXQEXXXKXEE 32  
DB 486 GEEEGAEAEAEAKDESDTKKEEGEGEE 517

RESULT 12  
TBFL\_YEAST  
ID TBFL\_YEAST STANDARD; PRT; 562 AA.  
AC Q02457;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE TBFL protein (TTAGGG repeat-binding factor 1) (TBF alpha).  
GN TBFL OR YPL128C OR LPI16C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LN224;  
RX MEDLINE=93140769; PubMed=8423796;  
RA Brigati C., Kurtz S., Balderes D., Vidali G., Shore D.M.;  
RT "An essential yeast gene encoding a TTAGGG repeat-binding protein.";  
RL Mol. Cell. Biol. 13:1306-1314(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / AB972;  
RX MEDLINE=97313271; PubMed=9169875;  
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoerge W.,  
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,  
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,  
RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,  
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,  
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,  
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,  
RA Scharens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,  
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Viessers S., Voss H.,  
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,  
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;  
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";  
RL Nature 387:103-105(1997).  
CC -!- FUNCTION: Binds the telomeric double-stranded TTAGGG repeat and  
CC negatively regulates telomere length. Involved in the regulation  
CC Essential for cell growth. May regulate gene expression.  
CC -!- SUBUNIT: Homodimer (Probable).  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X69394; CAA49191.1; -  
DR EMBL; U43703; AAB68230.1; -  
DR PIR; S28558; S28558.  
DR PIR; A48079; A48079.  
DR HSSP; P54274; 1BA5.  
DR TRANSFAC; T01246; -  
DR SGD; S0006049; TBFL.  
DR Pfam; PF00249; myb.DNA-binding; 1.  
DR SMART; SM00395; SANT; 1.  
DR PROSITE; PS00037; MYB\_1; 1.  
DR PROSITE; PS00090; MYB\_3; 1.  
KW Transcription regulation; Cell cycle; Nuclear protein;  
KW Chromosomal protein; Telomere; DNA-binding.  
FT DNA\_BIND 404 456 MYB.  
FT CONFLICT 71 71 E -> R (IN REF. 1).  
FT CONFLICT 371 373 DAA -> ERR (IN REF. 1).

SQ SEQUENCE 562 AA; 62823 MW; 2340F086468BC54F CRC64;

Query Match 30.6%; Score 33; DB 1; Length 562;  
Best Local Similarity 53.8%; Pred. No. 28;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 13 AXKXSXTQXQE 25  
| : |||||  
DB 516 AASATEDTQHQE 528

## RESULT 13

APM2\_YEAST  
ID APM2\_YEAST STANDARD; PRT; 605 AA.  
AC P38700;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Adaptin medium chain homolog APM2.  
GN APM2 OR YHL019C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / REE526;  
RA MEDLINE=95268148; PubMed=7749194;  
RA Lemmon S.K.;  
RA Stepp J.D., Pellucena-Palle A., Hamilton S., Kirchhausen T.,  
"A late Golgi sorting function for Saccharomyces cerevisiae Apm1p,  
but not for Apm2p, a second yeast clathrin AP medium chain-related  
protein.";  
RL Mol. Biol. Cell 6:41-58(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA MEDLINE=94378003; PubMed=8091229;  
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,  
Du Z., Favell A., Fulton L., Gattung S., Gelsel C., Kirsten J.,  
Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,  
Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,  
Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,  
Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,  
Vaudin M.;  
RA "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
VIII.";  
RL Science 265:2077-2082(1994).  
CC -1- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES MEDIUM SUBUNITS  
FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC EMBL: U11582; AAB65072.1; -  
CC DR PIR: S46833; S46833.  
CC DR SGD: S0001011; APM2.  
CC DR InterPro: IPR001392; Clathrn\_med.  
CC DR Pfam: PF00928; Adap\_comp\_sub; 1.  
CC DR PROSITE: PS00990; CLAT\_ADAPTOR\_M.1; 1.  
CC DR PROSITE: PS00991; CLAT\_ADAPTOR\_M.2; 1.  
CC Coated pits.  
CC -----

SQ SEQUENCE 605 AA; 69990 MW; 7E216B11325EEB3C CRC64;  
Query Match 30.6%; Score 33; DB 1; Length 605;  
Best Local Similarity 31.8%; Pred. No. 30;  
Matches 7; Conservative 2; Mismatches 13; Indels 0; Gaps 0;

QY 1 GHYKXNKNKSYXAXXKSDTQT 22  
| : : |||  
DB 476 GHREHSTNKSQYNSDEDDPNT 497

## RESULT 14

E2F\_DROME  
ID E2F\_DROME STANDARD; PRT; 805 AA.  
AC Q27368; O77035;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 13-JUN-2002 (Rel. 41, Last annotation update)  
DE Transcription factor E2f (de2f).  
GN E2F OR E2F1 OR CG6376.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Eye imaginal disk;  
RX MEDLINE=94294381; PubMed=8022787;  
RA Dynlacht B.D., Brook A., Dembski M., Yenush L., Dyson N.;  
RT "DNA-binding and trans-activation properties of Drosophila E2F and DP  
proteins.";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:6359-6363(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=94158833; PubMed=8114698;  
RA Ontani K., Nevins J.R.;  
RT "Functional properties of a Drosophila homolog of the E2f1 gene.";  
RL Mol. Cell. Biol. 14:1603-1612(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Oregon-R;  
RX MEDLINE=99077992; PubMed=9858578;  
RA Sasaki T., Sawado T., Yamaguchi M., Shimomiya T.;  
RT "Specification of regions of DNA replication initiation during  
RT embryogenesis in the 65-kilobase DNAlalpha-de2f locus of Drosophila  
RT melanogaster.";  
RL Mol. Cell. Biol. 19:547-555(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=107311132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Cocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abrah J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Botchan M.R., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botschan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foslodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrtka R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zeng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [5]  
RP FUNCTION.  
RX MEDLINE=97415323; PubMed=92711122;  
RA Royzman I., Whittaker A.J., Orr-Weaver T.L.;  
RT "Mutations in *Drosophila* Dp and E2F distinguish G1-S progression from  
RT an associated transcriptional program.";  
RL Genes Dev. 11:1999-2011(1997).  
RN [6]  
RP FUNCTION.  
RX MEDLINE=98078671; PubMed=9418862;  
RA Duronio R.J., Bonnette P.C., O'Farrell P.H.;  
RT "Mutations of the *Drosophila* dBP, dE2F, and cyclin E genes reveal  
RT distinct roles for the E2F-Dp transcription factor and cyclin E  
RT during the G1-S transition.";  
RL Mol. Cell. Biol. 18:141-151(1998).  
RN [7]  
RP TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RX MEDLINE=96087653; PubMed=8537434;  
RA Hao X.F., Alphey L., Bandara L.R., Lam E.W., Glover D.,  
RA La Thangue N.B.;  
RT "Functional conservation of the cell cycle-regulating transcription  
RT factor DRTF1/E2F and its pathway of control in *Drosophila*  
RT melanogaster.";  
RL J. Cell Sci. 108:2945-2954(1995).  
CC -!- FUNCTION: Transcriptional activator that binds to E2F sites.  
CC Required for wild-type growth in mitotic and polytene tissues,  
CC contributes to the expression of replication genes at the G1-S  
CC transition and Cyclin E.  
CC -!- SUBUNIT: HETERODIMER OF E2F AND DP. COOPERATE TO GIVE SEQUENCE-  
CC SPECIFIC DNA BINDING AND OPTIMAL TRANS-ACTIVATION.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- TISSUE SPECIFICITY: Segmentally repeated expression throughout  
CC early embryos is restricted to the ventral nerve cord in later  
CC embryos.  
CC -!- DEVELOPMENTAL STAGE: Throughout embryonic development.  
CC -!- SIMILARITY: BELONGS TO THE E2F/DP FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X78421; AAA55186.1; -  
CC EMBL; U10184; AAA19003.1; -  
CC EMBL; AB011813; BAA32746.1; -  
CC EMBL; AB003735; AAF55904.1; -  
CC HSSP; Q16234; ICF7.  
CC TRANSFAC; T01547; -  
CC FlyBase; FBgn0011766; E2F.  
CC InterPro; IPR003316; E2F\_TDP.  
CC Pfam; PF02319; E2F\_TDP; 1.  
KW Transcription regulation; DNA-binding; Nuclear protein;  
KW Developmental protein; Activator.  
FT DNA\_BIND 253 318 POTENTIAL.  
FT DOMAIN 318 411 DIMERIZATION (POTENTIAL).  
FT DOMAIN 14 19 POLY-SER.  
FT DOMAIN 64 68 POLY-ASN.  
FT DOMAIN 115 125 POLY-ALA.  
FT DOMAIN 129 143 GLY-RICH.

FT DOMAIN 245 249 POLY-SER.  
FT DOMAIN 519 573 GLY-RICH.  
FT DOMAIN 525 533 POLY-GLN.  
FT DOMAIN 594 601 POLY-ALA.  
FT DOMAIN 701 710 POLY-GLY.  
FT CONFLICT 127 127 H -> Q (IN REF. 1 AND 2).  
SQ SEQUENCE 805 AA; 87460 MW; BD49C813DB2A77D CRC64;  
Query Match 30.6%; Score 33; DB 1; Length 805;  
Best Local Similarity 35.0%; Pred. No. 41;  
Matches 7; Conservative 2; Mismatches 11; Indels 0; Gaps 0;  
QY 6 NXNKSXXAXKXSDTQTQXOE 25  
DB 13 NSNSSSHPTTSSNTQRHQ 32  
RESULT 15  
V33P\_ADE41 STANDARD; PRT; 217 AA.  
ID V33P\_ADE41  
AC P19416; O64818;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 01-MAR-1992 (Rel. 21, Last annotation update)  
DE 33 kDa phosphoprotein.  
OS Human adenovirus type 41.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=10524;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Tak;  
RX MEDLINE=90272433; PubMed=2349115;  
RA Slemenda S.B., Pieniazek N.J., Velarde J. Jr., Pieniazek D.,  
RA Luftig R.B.;  
RT "Nucleotide sequence of the region coding for 100K and 33K proteins  
RT of human enteric adenovirus type 41 ('Tak').";  
RL Nucleic Acids Res. 18:3069-3069(1990).  
RN [2]  
RP SEQUENCE OF 203-217 FROM N.A.  
RC STRAIN=Tak;  
RX MEDLINE=89345113; PubMed=2762136;  
RA Pieniazek N.J., Velarde J. Jr., Pieniazek D., Luftig R.B.;  
RT "Nucleotide sequence of human enteric adenovirus type 41 hexon-  
RT associated protein VIII precursor (pVIII) including the early region  
RT E3 promoter.";  
RL Nucleic Acids Res. 17:5398-5398(1989).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X52532; CAB38632.1; -  
CC EMBL; X15137; CAA33236.1; -  
CC PIR; S04851; S04851.  
CC PIR; S10212; S10212.  
KW Late protein; Phosphorylation.  
SQ SEQUENCE 217 AA; 24740 MW; B6ECDAB3C24EC3E CRC64;  
Query Match 29.6%; Score 32; DB 1; Length 217;  
Best Local Similarity 36.0%; Pred. No. 17;  
Matches 9; Conservative 1; Mismatches 15; Indels 0; Gaps 0;  
QY 8 NKSSXXAXKXSDTQTQXEXXEE 32  
DB 6 NKQATADRSQKQKLEQWDEEE 30  
Search completed: July 16, 2003, 17:47:28  
Job time : 3.8887 secs



```
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE RE17222p.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY071126; AAL48748.1; -.
SQ SEQUENCE 198 AA; 22012 MW; CE33F831FAB79D49 CRC64;

Query Match 35.3%; Score 49; DB 5; Length 198;
Best Local Similarity 42.9%; Pred. No. 35;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Qy 3 PNTNKEKAERKSHDTQTQE 23
Db 118 PNRNEEQATRRSESTSTQAE 138

RESULT 15
Q92951
ID Q92951 PRELIMINARY; PRT; 273 AA.
AC Q92951;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Zinc finger protein (Hypothetical 31.0 kDa protein).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=98135459; Pubmed=9475118;
RA Drew P.D., Gado A.M., Canning R.D., Nagle J.W., Dehejia A.M.,
RA Polymeropoulos M.H., Biddison W.E., Jacobson S., Becker K.G.;
RT "C2H2-546: a zinc finger protein differentially expressed in HTLV-1
RT infected T cells.";
RL J. Neurovirol. 3:455-459(1997).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=BRAIN;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; U69645; AAB09561.1; -.
DR EMBL; BC022842; AAH22842.1; -.
DR HSSP; P08046; 1A1H.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 7.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 4.
DR SMART; SM00355; Znf_C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
KW Hypothetical protein; DNA-binding; Metal-binding; Nuclear protein;
KW Zinc-finger.
SQ SEQUENCE 273 AA; 31029 MW; 91E7148C665C4734 CRC64;

Query Match 35.3%; Score 49; DB 4; Length 273;
Best Local Similarity 50.0%; Pred. No. 48;
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy 4 NTKNKEKAERKSHDTQTQE 23
Db 48 NSFREKLEKSPDSKTLQE 67

Search completed: July 16, 2003, 17:50:05
Job time : 7.31366 secs
```

RESULT 12	
Q9PM90	
ID	Q9PM90 PRELIMINARY; PRT; 923 AA.
AC	Q9PM90;
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Putative oxidoreductase.
DE	Putative oxidoreductase.
GN	CJ1585C.
OS	Campylobacter jejuni.
OC	Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC	Campylobacter.
OX	NCBI_TaxID=197;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NCTC 11168;
RC	MEDLINE=20150912; PubMed=10688204;
RA	Farkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA	Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA	Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA	Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA	Whitehead S., Barrell B.G.;
RT	"The genome sequence of the food-borne pathogen Campylobacter jejuni
RT	reveals hypervariable sequences.";
RL	Nature 403:665-668(2000).
RL	EMBL; AL139079; CAB73573.1; "
DR	InterPro: IPR001450; 4Fe4S_ferredoxin.
DR	InterPro: IPR004113; FAD-oxidase_C.
DR	InterPro: IPR001573; Oxid_FAD_bind.
DR	InterPro: IPR001467; Prok_Boxed.
DR	Pfam: PF02913; FAD-oxidase_C 1.
DR	Pfam: PF01565; FAD binding 4; 1.

```

RA Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RA SEQUENCE FROM N.A.
RA Jia Y., Cherry J.M.;
RA Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RA SEQUENCE OF 2156-3268 FROM N.A.
RP Utsugi T., Kikuchi Y.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U33050; AAB64910.1; -.
DR EMBL; D63905; BAA21482.1; -.
DR SGD; S0002865; TOM1.
DR InterPro; IPR000569; HECT_domain.
DR Pfam; PF00632; HECT; 1.
DR SMART; SM00119; HECTc; 1.
DR PROSITE; PS00237; HECT; 1.
KW Ligase.
SQ SEQUENCE 3268 AA; 374180 MW; 8F71F3493D70A6C9 CRC64;

Query Match 36.0%; Score 50; DB 3; Length 3268;
Best Local Similarity 44.0%; Pred. No. 4.3e+02;
Matches 11; Conservative 6; Mismatches 6; Indels 2; Gaps 1;

QY 2 KPNTNKSEKAERKSHDTQTTOEICE 26
Db 226 KNNINETQTSIKKV--TMTTQELCE 248
| | | | | : : : | | | | | | | |
| | | | | : : | | | | | | | |

RESULT 14
Q8SZ45 PRELIMINARY; PRT; 198 AA.
ID Q8SZ45
AC Q8SZ45;
DT 01-JUN-2002 (TEMBLrel. 21, Created).

```

```

QY 2 KPNTNK---SEKAERKSHDTQTTOEI 24
  || | |||| | : : |||
Db 131 KPEVKEVEGEKAEKEETEAETQEL 156

RESULT 8
ID O01439 PRELIMINARY; PRT; 196 AA.
AC O01439;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Hypothetical 21.5 kDa protein.
GN B0261.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None.
RT "Genome sequence of the nematode C. elegans: a platform for
  investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Scheet P., Gattung S.;
RT "The sequence of C. elegans cosmid B0261.";
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RD EMBL; U97016; AAB52352.1; -.
KW Hypothetical protein.
SQ SEQUENCE 196 AA; 21516 MW; AF8FFBE84C60EF48 CRC64;

Query Match 36.0%; Score 50; DB 5; Length 196;
Best Local Similarity 43.5%; Pred. No. 25;
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 KPNTNKSEKAERKSHDTQTTOEI 24
  || : ||| || ||||
Db 105 KADKRRKKKANKSPRTQEQEV 127

RESULT 9
Q56267 PRELIMINARY; PRT; 465 AA.
AC Q56267;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Glutamate synthase small subunit gltD.
GN GLTD.
OS Thioobacillus ferrooxidans.
OC Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.
OX NCBI_TaxID=920;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC33020;
RA Deane S.M., Rawlings D.E.;
RT "The gene for the Thioobacillus ferrooxidans ATCC33020 glutamate-
  synthase (GOGAT) small subunit complements an Escherichia coli gltD
  mutant.";
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -I- COFACTOR: FAD (BY SIMILARITY).
RD EMBL; U36427; AAA79783.1; -.

```

```

DR InterPro; IPR000759; Adrndx_reductase.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF00070; pyr_redox; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR TIGRFAMS; TIGR01318; gltd_gamma_fam; 1.
KW FAD; Flavoprotein; Oxidoreductase.
SQ SEQUENCE 465 AA; 50793 MW; B38EAD4B4D0BEE6A CRC64;

Query Match 36.0%; Score 50; DB 2; Length 465;
Best Local Similarity 50.0%; Pred. No. 59;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 6 NKSEKAERKSHDTQTTOEI 25
  | : | | | | | | | |
Db 73 NRLEEATLSHQINTLPEIC 92

RESULT 10
Q82S65 PRELIMINARY; PRT; 471 AA.
ID Q82S65;
AC Q82S65;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein ALR7647.
GN ALR7647.
OS Anabaena sp. (strain PCC 7120).
OC Plasmid pCC7120beta.
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
  cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
RD EMBL; AP003602; BAB77290.1; -.
DR InterPro; IPR005094; Relaxase.
DR Pfam; PF03432; Relaxase; 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 471 AA; 53809 MW; 7DDBB895E2209C7E CRC64;

Query Match 36.0%; Score 50; DB 16; Length 471;
Best Local Similarity 45.5%; Pred. No. 60;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 3 PNTNKSEKAERKSHDTQTTOEI 24
  || : ||| | | | | | | |
Db 151 PNRKSKIRAVNSKRKHTPTSKEI 172

RESULT 11
Q928K7 PRELIMINARY; PRT; 579 AA.
ID Q928K7;
AC Q928K7;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CT082 hypothetical protein.
GN CPN0331 OR CPJ0331 OR CP0426.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CW1029;
RX MEDLINE=99206606; PubMed=10192388;

```

```

DR Pfam; PF00618; RasGEFN; 1.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEF; 1.
DR PROSITE; PS00720; GDS_CDC25; 1.
SQ SEQUENCE 1048 AA; 121960 MW; 40A1F12F9B86F458 CRC64;

Query Match
Best Local Similarity 38.8%; Score 54; DB 3; Length 1048;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 3 PNTNKSEKAERKSHDTQTQ 22
DB 1002 PNSNSNNKSQEKSRDDOTDE 1021

RESULT 5
Q9JG50
AC Q9JG50 PRELIMINARY; PRT; 661 AA.
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ORF1.
OS TTV-like mini virus.
OC Viruses; ssDNA viruses; Circoviridae.
OX NCBI_TaxID=93678;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLMV-NLC026;
RA Mishi S., Hijikata M., Takahashi K.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TLMV-NLC026;
RX MEDLINE=20428649; PubMed=10971131;
RA Takahashi K., Hijikata M., Samokhvalov E.I., Mishi S.;
RT "Full or near full length nucleotide sequences of TTV virus variants (types SANBAN and YONBAN) and the TTV virus-like mini virus.";
RL Intervirology 43:119-123(2000).
DR EMBL; AB038630; BAA93609.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TTV_ORF1; 1.
SQ SEQUENCE 661 AA; 78995 MW; 83CE5F93A1AD286A CRC64;

Query Match
Best Local Similarity 51.7%; Score 52.5; DB 12; Length 661;
Matches 15; Conservative 3; Mismatches 6; Indels 5; Gaps 2;

QY 3 PNTNKSEKAERKSHDTQTQ-ICE 26
DB 566 PNTNKSEVLYCFDERRQLTKTAQRICE 594

RESULT 6
Q9Z7W9
AC Q9Z7W9 PRELIMINARY; PRT; 651 AA.
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SIMILARITY TO CHLPS INCA (CPJ0585 protein).
GN CPN0585 OR CPJ0585 OR CP0163.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=833558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).

```

---

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickie E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson M., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AE001643; AAD18724.1; -.
DR EMBL; AE002177; AAF38042.1; -.
DR EMBL; AP002547; BAA98792.1; -.
DR TIGR; CP0163; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 651 AA; 75668 MW; 7EAC83171535CAB8 CRC64;

Query Match
Best Local Similarity 36.7%; Score 51; DB 16; Length 651;
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 6 NKSEKAERKSHDTQTQ-ICE 26
DB 270 NQSKDIQRAHREASQACE 290

RESULT 7
Q8X093
AC Q8X093 PRELIMINARY; PRT; 642 AA.
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Related to endo-exonuclease yNucR.
GN B14D6.350.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356173; CAB91744.2; -.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR002792; TRAM.
DR InterPro; IPR001566; TrmA.
DR Pfam; PF01938; TRAM; 1.
DR PROSITE; PS01230; TRMA_1; UNKNOWN_1.
DR PROSITE; PS01231; TRMA_2; UNKNOWN_1.
KW Exonuclease.
SQ SEQUENCE 642 AA; 70860 MW; F064B8462E44BB43 CRC64;

Query Match
Best Local Similarity 36.3%; Score 50.5; DB 3; Length 642;
Matches 11; Conservative 4; Mismatches 8; Indels 3; Gaps 1;

```



```

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Seroreactive antigen BMN1-17 (Fragment).
OS Babesia microti.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
OX NCBI_TaxID=5868;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MNI;
RX MEDLINE=20231818; PubMed=10768973;
RA Lodes M.J., Houghton R.L., Bruinsma E.S., Mohamath R., Reynolds L.D.,
RA Benson D.R., Krause P.J., Reed S.G., Persing D.H.;
RT "Serological expression cloning of novel immunoreactive antigens of
RT Babesia microti.";
RL Infect. Immun. 68:2783-2790(2000).
DR EMBL; AF206526; AAF68253.1; -.
FT NON_TER 396 396
SQ SEQUENCE 396 AA; 44720 MW; 706E153BA5EB7C CRC64;

Query Match 100.0%; Score 139; DB 5; Length 396;
Best Local Similarity 100.0%; Pred. No. 2.1e-11;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPTNKSEAKERSHDTQTTOICE 26
Db 133 GKPTNKSEAKERSHDTQTTOICE 158

RESULT 3
Q20227 PRELIMINARY; PRT; 879 AA.
AC Q20227;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F40F8.5 protein.
GN F40F8.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA MacDougall R.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=95069613; PubMed=9851916;
RT none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z69302; CAA93261.1; -.
DR InterPro; IPR003439; ABC_transport.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
SQ SEQUENCE 879 AA; 99694 MW; BE4403D836B7E3C2 CRC64;

Query Match 43.5%; Score 60.5; DB 5; Length 879;
Best Local Similarity 52.2%; Pred. No. 3.9;
Matches 12; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Qy 5 TNKSEK-AERKSHDTQTTOICE 26
Db 3 TNKREQLERECHQKTSQAVCE 25

RESULT 4
Q12037 PRELIMINARY; PRT; 1048 AA.
ID Q12037;
AC Q12037;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

```

```

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein I1309.
GN SDC25 OR SCD25 OR L1309/SCD25 OR YL1016W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ALPHA S288C;
RA Purnelle B., Goffeau A.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Goffeau A., Purnelle B.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Miosga T., Zimmermann F.K.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=FY23 /RD005;
RX MEDLINE=96405918; PubMed=8810043;
RA Miosga T., Zimmermann F.K.;
RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT a 43.7 kb fragment of chromosome XII including an open reading frame
RT regulator protein CFTF.";
RL Yeast 12:693-708(1996).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=FY23 /RD005;
RX MEDLINE=91094833; PubMed=1986220;
RA Damak F., Boy-Marcotte E., Le-Roscouet D., Guilbaud R., Jacquet M.;
RT "SDC25, a CDC25-like gene which contains a RAS-activating domain and
RT is a dispensable gene of Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 11:202-212(1991).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=FY23 /RD005;
RX MEDLINE=93087480; PubMed=1454790;
RA Ripmaster T.L., Vaughn G.P., Woolford J.L. Jr.;
RT "A putative ATP-dependent RNA helicase involved in Saccharomyces
RT cerevisiae ribosome assembly.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135(1992).
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=FY23 /RD005;
RX MEDLINE=94375516; PubMed=8089172;
RA Burgess S.M., Delannoy M., Jensen R.E.;
RT "MMI encodes a mitochondrial outer membrane protein essential for
RT establishing and maintaining the structure of yeast mitochondria.";
RL J. Cell Biol. 126:1375-1391(1994).
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=FY23 /RD005;
RX MEDLINE=95348179; PubMed=7622557;
RA Gammie A.E., Kurihara L.J., Vallee R.B., Rose M.D.;
RT "DNM1, a dynamin-related gene, participates in endosomal trafficking
RT in yeast.";
RL J. Cell Biol. 130:553-566(1995).
DR EMBL; X97560; CAA66161.1; -.
DR EMBL; Z73121; CAA97461.1; -.
DR EMBL; X91488; CAA62775.1; -.
DR SGD; S0003939; SDC25.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGEF.
DR Pfam; PF00617; RasGEF; 1.

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:37:48 ; Search time 6.31366 seconds  
(without alignments)  
848.513 Million cell updates/sec

Title: US-09-853-079-195  
Perfect score: 139  
Sequence: 1 GKPNINKSEAKRSHDTQTTOICE 26

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	139	100.0	275	5 Q9NIM2	Q9nim2 babesia mic
2	139	100.0	396	5 Q9NIM4	Q9nim4 babesia mic
3	60.5	43.5	879	5 Q20227	Q20227 caenorhabdi
4	54	38.8	1048	3 Q12037	Q12037 saccharomyc
5	52.5	37.8	661	12 Q9JG50	Q9jg50 tiv-like mi
6	51	36.7	651	16 Q9Z7W9	Q9z7w9 chlamydia p
7	50.5	36.3	642	3 Q8X093	Q8x093 neurospora
8	50	36.0	196	5 Q01439	Q01439 caenorhabdi
9	50	36.0	465	2 Q56267	Q56267 thioobacillu
10	50	36.0	471	16 Q8Z865	Q8z865 anabaena sp
11	50	36.0	579	16 Q9Z8K7	Q9z8k7 chlamydia p
12	50	36.0	923	16 Q9PW90	Q9pm90 campylobact
13	50	36.0	3268	3 Q83280	Q83280 saccharomyc
14	49	35.3	198	5 Q8S245	Q8sz45 drosophila
15	49	35.3	273	4 Q9Z951	Q9z951 homo sapien
16	49	35.3	400	5 Q9VB99	Q9vby9 drosophila

17	49	35.3	803	3	013657	013657 schizosacch
18	48.5	34.9	529	10	Q9FN33	Q9fn33 arabidopsis
19	48.5	34.9	733	5	Q8SWD6	Q8swd6 encephalito
20	48.5	34.9	5038	11	Q9QYX7	Q9qyx7 mus musculu
21	48	34.5	149	10	Q8WI91	Q8wi91 arabidopsis
22	48	34.5	506	10	Q9LUN6	Q9lun6 arabidopsis
23	48	34.5	579	5	076734	076734 dictyosteli
24	47.5	34.2	211	16	031611	031611 bacillus su
25	47.5	34.2	474	5	Q9VWQ6	Q9vwq6 drosophila
26	47.5	34.2	474	5	Q9VWQ3	Q9vwq3 drosophila
27	47	33.8	48	11	Q9RLY9	Q9rly9 mus musculu
28	47	33.8	89	5	Q9BJN1	Q9bjn1 plasmodium
29	47	33.8	92	5	Q9BMT2	Q9bmt2 plasmodium
30	47	33.8	99	4	Q9BZT7	Q9bzt7 homo sapien
31	47	33.8	208	10	Q93YZ3	Q93yz3 arabidopsis
32	47	33.8	220	11	Q9Z2V0	Q9z2v0 mus musculu
33	47	33.8	332	11	Q9QW73	Q9qw73 mus sp. bet
34	47	33.8	453	16	083072	083072 treponema p
35	47	33.8	537	10	Q9C551	Q9c551 arabidopsis
36	47	33.8	635	5	096137	096137 plasmodium
37	47	33.8	753	5	Q97294	Q97294 plasmodium
38	47	33.8	939	5	015790	015790 plasmodium
39	47	33.8	939	5	015799	015799 plasmodium
40	47	33.8	939	5	Q9NG57	Q9ng57 plasmodium
41	47	33.8	950	5	015788	015788 plasmodium
42	47	33.8	1008	10	Q9FNE4	Q9fne4 arabidopsis
43	47	33.8	1097	10	Q9SD87	Q9sd87 arabidopsis
44	47	33.8	2154	11	Q9QWJ7	Q9qwj7 mus musculu
45	46.5	33.5	65	16	Q9PMJ6	Q9pmj6 campylobact

## ALIGNMENTS

## RESULT 1

Q9NIM2 ID Q9NIM2 PRELIMINARY; PRT; 275 AA.  
AC Q9NIM2:  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE Seroreactive antigen BMN1-20 precursor (Fragment).  
OS Babesia microti.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
OX NCBI\_TaxID=5868;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MN1;  
RX MEDLINE=20231818; PubMed=10768973;  
RA Lodes M.J., Houghton R.L., Bruinsma E.S., Mohamath R., Reynolds L.D.,  
RA Benson D.R., Krause P.J., Reed S.G., Persing D.H.;  
RT "Serological expression cloning of novel immunoreactive antigens of  
RT Babesia microti.";  
RL Infect. Immun. 68:2783-2790(2000).  
DR EMBL; AF206527; AAF68255.1; -;  
KW Signal.  
FT SIGNAL 1 22 POTENTIAL.  
FT NON\_TER 275  
SQ SEQUENCE 275 AA; 30775 MW; C3CC5D5EBC494E01 CRC64;

Query Match 100.0%; Score 139; DB 5; Length 275;  
Best Local Similarity 100.0%; Pred. No. 1.4e-11;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GKPNINKSEAKRSHDTQTTOICE 26  
|||||  
Db 218 GKPNINKSEAKRSHDTQTTOICE 243

## RESULT 2

Q9NIM4 ID Q9NIM4 PRELIMINARY; PRT; 396 AA.  
AC Q9NIM4;

```
-----
CC EMBL; Z81035; CAB02731.1; -.
DR WormPep; C15H11.9; CE08186.
KW Ribosome biogenesis; Nuclear protein.
FT DOMAIN 79 332 LYS-RICH.
SQ SEQUENCE 333 AA; 37985 MW; 88E91737402FC4CE CRC64;

Query Match 32.4%; Score 45; DB 1; Length 333;
Best Local Similarity 47.6%; Pred. No. 43;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 3 PNTNKSKEAERKSHDTQTQOE 23
   | | | | | | | | | |
DB 277 PLREKKEKSERKAGKQDTROK 297

RESULT 15
YC00_MYCPN STANDARD; PRT; 798 AA.
AC Q50286;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein MPN200 precursor (GT9_orf798).
GN MPN200 OR MP631.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=96177562; PubMed=8604303;
RA Hilbert H., Himmelreich R., Plagens H., Herrmann R.;
RT "Sequence analysis of 56 kb from the genome of the bacterium
   Mycoplasma pneumoniae comprising the dnaA region, the atp operon and a
   cluster of ribosomal protein genes.";
RL Nucleic Acids Res. 24:628-639(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
   Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
   pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
   (Potential).
CC -!- SIMILARITY: BELONGS TO THE MG185 / MG260 FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; U34795; AAC43681.1; -.
DR EMBL; AE000060; AAB96279.1; -.
DR InterPro; IPR004890; Lipoprotein_10.
DR InterPro; IPR004984; Lipoprotein_X.
DR Pfam; PF03202; Lipoprotein_10; 1.
DR Pfam; PF03305; Lipoprotein_X; 1.
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 798 HYPOTHETICAL LIPOPROTEIN MPN200.
FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 798 AA; 87628 MW; 3E2471D7EFOA3CA0 CRC64;

Query Match 32.4%; Score 45; DB 1; Length 798;
Best Local Similarity 40.9%; Pred. No. 1e+02;
-----
```

```
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 KPNTNKSKEAERKSHDTQTQOE 23
   | | | | | | | | | |
DB 229 KKEKKSTRSKRATEGTQTKE 250

Search completed: July 16, 2003, 17:47:32
Job time : 3.53457 secs
```









RT "The COOH-domain of the product of the Saccharomyces cerevisiae SCD25  
 RL gene elicits activation of p21-ras proteins in mammalian cells."  
 CC -|- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.  
 CC -|- MISCELLANEOUS: SUPPRESSES THE CDC25-5 MUTATION IN YEAST (RESTORES  
 CC CAMP LEVEL) AND HAS SIMILAR FUNCTIONS AS CDC25.  
 CC -|- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -|- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M26647; AA16565.1; -  
 DR PIR; S14177; S14177.  
 DR SGD; S0003939; SDC25.  
 DR InterPro; IPR000651; RasGEF.  
 DR InterPro; IPR001895; RasGEF\_CDC25.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00018; SH3; 1.  
 DR Pfam; PF00617; RasGEF; 1.  
 DR Pfam; PF00618; RasGEF; 1.  
 DR SMART; SM00147; RasGEF; 1.  
 DR SMART; SM00229; RasGEF; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS00720; GDS\_CDC25; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW Guanine-nucleotide releasing factor; Cell cycle; Cell division;  
 KW SH3 domain.  
 FT DOMAIN 26 98 SH3.  
 FT DOMAIN 74 79 POLY-ASN.  
 FT DOMAIN 434 437 POLY-ARG.  
 FT VARIANT 584 590 DVVVF -> V (IN STRAIN OLL136).  
 SQ SEQUENCE 1253 AA; 144979 MW; 2DE2C9EC27E3E60D CRC64;  
 Query Match 38.8%; Score 54; DB 1; Length 1253;  
 Best Local Similarity 45.0%; Pred. No. 9.8;  
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
 QY 3 PNTNKEAERKSHDTQTQ 22  
 DB 1207 PMSNSNKSQKSRDDQTDE 1226  
 RESULT 4  
 ID AEX3 CAEEL STANDARD; PRT; 1409 AA.  
 AC Q02626; Q27467;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Mitochondrial import receptor subunit TOM20-2 (Translocase of outer  
 DE membrane 20 kDa subunit 2).  
 GN AEX-3 OR C02H7.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=97282461; PubMed=9136770;  
 RA Iwasaki K., Staunton J., Saifee O., Nonet M., Thomas J.H.;  
 RT "aex-3 encodes a novel regulator of presynaptic activity in C.  
 RT elegans."  
 RL Neuron 18:613-622(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Leimbac D., Minx M.;

Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RL INTERACTION WITH CAB-1.  
 RN MEDLINE=20428446; PubMed=10970871;  
 RX Iwasaki K., Toyonaga R.;  
 RA "The rab3 GDP/GTP exchange factor homolog AEX-3 has a dual function  
 RT in synaptic transmission."  
 RT EMBO J. 19:4806-4816(2000).  
 RL -|- FUNCTION: GUANINE NUCLEOTIDE EXCHANGE FACTOR (GEF) FOR RAB3. MAY  
 CC REGULATE TWO DIFFERENT PATHWAYS FOR NEURAL ACTIVITIES.  
 CC -|- SUBUNIT: BINDS TO CAB-1.  
 CC -|- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF NEURONS.  
 CC -|- SIMILARITY: CONTAINS 1 DENN DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U93842; AAB52421.1; -  
 DR EMBL; U49945; AAC47926.1; -  
 DR WormPep; C02H7.3; CE16806.  
 DR InterPro; IPR001194; DENN.  
 DR InterPro; IPR005112; dDENN.  
 DR InterPro; IPR005113; uDENN.  
 DR Pfam; PF02141; DENN; 1.  
 DR Pfam; PF03455; dDENN; 1.  
 DR Pfam; PF03456; uDENN; 1.  
 KW Guanine-nucleotide releasing factor.  
 FT DOMAIN 216 364 DENN.  
 FT DOMAIN 604 632 SER-RICH.  
 SQ SEQUENCE 1409 AA; 157458 MW; 2DDE6395AC963313 CRC64;  
 Query Match 36.0%; Score 50; DB 1; Length 1409;  
 Best Local Similarity 32.0%; Pred. No. 39;  
 Matches 8; Conservative 7; Mismatches 10; Indels 0; Gaps 0;  
 QY 2 KPNTNKEAERKSHDTQTQICE 26  
 DB 769 QPNPTSQQTANQSKNQQTVEKFD 793  
 RESULT 5  
 ID OM22 ARATH STANDARD; PRT; 210 AA.  
 AC P82873; Q9FZJ6;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Mitochondrial import receptor subunit TOM20-2 (Translocase of outer  
 DE membrane 20 kDa subunit 2).  
 GN TOM20-2 OR AT1G27390 OR F17L21.18.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 37-48; 105-114 AND 132-147.  
 RC STRAIN=cv. Columbia;  
 RX PubMed=11161051;  
 RA Werhahn W., Niemeyer A., Jaensch L., Kruff V., Schmitz U.K.,  
 RA Braun H.-P.;  
 RT "Purification and characterization of the preprotein translocase of  
 RT the outer mitochondrial membrane from Arabidopsis thaliana.  
 RT Identification of multiple forms of TOM20.";  
 RL Plant Physiol. 125:943-954(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;



```
DR InterPro; IPR001305; DnaJ_CXXCXGXG.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ_1.
DR Pfam; PF00684; DnaJ_CXXCXGXGXG; 1.
DR Pfam; PF01556; DnaJ_C; 1.
DR PRINTS; PR00825; DnaJPROTEIN.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS00637; DnaJ_CXXCXGXGXG; 1.
DR PROSITE; PS00076; DnaJ_2; 1.
KW Chaperone; Heat shock; Mitochondrion; Repeat; Transit peptide.
FT TRANSIT 1 55 MITOCHONDRION (POTENTIAL).
FT CHAIN 56 511 MDJ1 PROTEIN.
FT DOMAIN 59 127 J-DOMAIN.
FT REPEAT 130 186 GLY-RICH.
FT REPEAT 230 237 CXXCXGXG MOTIF.
FT REPEAT 247 254 CXXCXGXG MOTIF.
FT REPEAT 269 276 CXXCXGXG MOTIF.
FT REPEAT 285 292 CXXCXGXG MOTIF.
SQ SEQUENCE 511 AA; 55561 MW; 007343427C66D2B6 CRC64;

Query Match 40.3%; Score 56; DB 1; Length 511;
Best Local Similarity 50.08; Pred. No. 2.1;
Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 3 PNTNKEKAERKSHDTQTTOEI 24
I: || || || || || || ||
Db 90 PDINKEPDAEKKFHDQLNAVET 111

RESULT 2
R114_HUMAN
ID R114_HUMAN STANDARD; PRT; 1158 AA.
AC P48552;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nuclear factor R1P140 (Nuclear receptor interacting protein 1).
GN NR1P1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast;
RX MEDLINE=95369246; PubMed=7641693;
RA Cavaillès V., Dauvois S., L'Horsset F., Lopez G., Hoare S.,
RA Kushner P.J., Parker M.G.;
RT "Nuclear factor R1P140 modulates transcriptional activation by the
RT estrogen receptor.";
RL EMBO J. 14:3741-3751(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordisiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
RA Ranser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gaspar M.-L.;
RA Lehnach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
RL Nature 405:311-319(2000).
CC -!- FUNCTION: MODULATES TRANSCRIPTIONAL ACTIVATION BY THE ESTROGEN
CC RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Nuclear.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X84373; CAA59108.1; -.
DR EMBL; AF248484; AAF62185.1; -.
DR EMBL; AF127577; AAF35255.1; -.
DR EMBL; AL163207; CAB90396.1; -.
DR GenBank; HGNC:8001; NR1P1.
DR MIM; 602490; -.
KW Transcription regulation; Nuclear protein.
FT CONFLICT 124 124 P -> R (IN REF. 1).
FT CONFLICT 448 448 R -> G (IN REF. 1).
FT CONFLICT 721 726 NKGSSE -> TKGRVK (IN REF. 1).
FT CONFLICT 1080 1080 T -> A (IN REF. 1).
SQ SEQUENCE 1158 AA; 126941 MW; 81FC424968E9A5F6 CRC64;

Query Match 38.8%; Score 54; DB 1; Length 1158;
Best Local Similarity 46.28; Pred. No. 9;
Matches 12; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1 GKPNTNKEKAERKSHDTQTTOEICE 26
I: || || || || || || ||
Db 718 GNPKNKGSKKERTPLRDESTQEHSE 743

RESULT 3
SC25_YEAST
ID SC25_YEAST STANDARD; PRT; 1253 AA.
AC P1471;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SCD25 protein.
GN SCD25 OR SDC25 OR YLL016W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=W303;
RX MEDLINE=91094833; PubMed=1986220;
RA Damak F., Boy-Marcotte E., le Roscouet D., Guilbaud R., Jacquet M.;
RT "SCD25, a CDC25-like gene which contains a RAS-activating domain and
RT is a dispensable gene of Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 11:202-212(1991).
RN [2]
RP SEQUENCE OF 668-1253 FROM N.A.
RC STRAIN=OHL136;
RX MEDLINE=89306677; PubMed=2545538;
RA Boy-Marcotte E., Damak F., Camonis J., Garreau H., Jacquet M.;
RT "The C-terminal part of a gene partially homologous to CDC 25 gene
RT suppresses the cdc25-5 mutation in Saccharomyces cerevisiae.";
RL Gene 77:21-30(1989).
RN [3]
RP FUNCTION.
RX MEDLINE=90206633; PubMed=2188363;
RA Crechet J.B., Pouillet P., Mistou M.-Y., Parmeggiani A., Camonis J.,
RA Boy-Marcotte E., Damak F., Jacquet M.;
RT "Enhancement of the GDP-GTP exchange of RAS proteins by the carboxyl-
RT terminal domain of SCD25.";
RL Science 248:866-868(1990).
RN [4]
RP FUNCTION.
RX MEDLINE=91156312; PubMed=2000228;
RA Rey I., Schweighoffer F., Barlat I., Camonis J., Boy-Marcotte E.,
RA Guilbaud R., Jacquet M., Tocque B.;
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:37:24 ; Search time 1.53457 Seconds  
(without alignments)  
702.728 Million cell updates/sec

Title: US-09-853-079-195

Perfect score: 139

Sequence: 1 GKPNTNKSEKAERKSHDTQTQICE 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	56	40.3	511	MDJ1_YEAST	P35191 saccharomyc
2	54	38.8	1158	R114_HUMAN	P48552 homo sapien
3	54	38.8	1253	SC25_YEAST	P14771 saccharomyc
4	50	36.0	1409	AXE3_CAEEL	O02626 caenorhabdi
5	49	35.3	210	OM22_ARATH	P82873 arabidopsis
6	47	33.8	120	TR17_HUMAN	Q35651 homo sapien
7	47	33.8	178	FY15_HUMAN	Q96db9 homo sapien
8	47	33.8	369	VP6_AHSV3	Q64909 african hor
9	47	33.8	369	VP6_AHSV6	Q64913 african hor
10	47	33.8	2363	SPCO_MOUSE	O62261 mus musculu
11	46	33.1	533	YN34_YEAST	P48565 saccharomyc
12	45	32.4	148	CALM_EUGGR	P11118 euglena gra
13	45	32.4	263	SGS3_DROVA	P13728 drosophila
14	45	32.4	333	RNS1_CAEEL	Q9xvt0 caenorhabdi
15	45	32.4	798	YC00_MYCPN	Q50288 mycoplasma
16	45	32.4	843	YMS1_YEAST	Q05050 saccharomyc
17	45	32.4	929	R5M4_HUMAN	P98175 homo sapien
18	45	32.4	1020	NFH_HUMAN	P12036 homo sapien
19	44.5	32.0	906	CENC_MOUSE	P49452 mus musculu
20	44	31.7	60	RL32_DETRA	P49228 deinococcus
21	44	31.7	327	HUNB_MANSE	Q25514 manduca sex
22	44	31.7	328	VP6_BTIVIS	Q03328 blueetongue
23	44	31.7	343	SLAM_MOUSE	Q9qum4 mus musculu
24	44	31.7	607	PESC_SCHPO	O60164 schizosacch
25	44	31.7	1803	YUL3_YEAST	P47024 saccharomyc
26	43.5	31.3	178	YK01_ARCFU	O28278 archaeoglob
27	43	30.9	52	RK32_CYACA	O19894 cyanidium c
28	43	30.9	56	RL32_MYCPN	P75238 mycoplasma
29	43	30.9	63	RL32_XYLFA	Q9pcq5 xylella fas
30	43	30.9	325	LBX1_PHOLE	P09141 photobacter
31	43	30.9	388	CHF2_SCHPO	P87125 schizosacch
32	43	30.9	406	SNX6_HUMAN	Q9unh7 homo sapien
33	43	30.9	436	Y868_CHLMU	Q9pfj91 chlamydia m

#### RESULT 1

ID	MDJ1_YEAST	STANDARD;	PRT;	511 AA.
AC	P35191;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	MDJ1 protein, mitochondrial precursor.			
GN	MDJ1 OR YFL016C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-S288C / AB972;			
RA	Rowley N.K., Prip-Buus C., Westermann B., Brown C.M., Schwarz E.,			
RA	Barrell B.G., Neupert W.;			
RL	Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-S288C / AB972;			
RA	Barrell B.G., Churcher C., Rajandream M.A.;			
RL	Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-S288C / AB972;			
RA	Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,			
RA	Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,			
RA	Yamazaki M., Tashiro H., Eki T.;			
RL	"Analysis of the nucleotide sequence of chromosome VI from			
RL	Saccharomyces cerevisiae".			
RT	Nat. Genet. 10:261-268(1995).			
CC	-1- FUNCTION: PLAYS A ROLE IN MITOCHONDRIAL BIOGENESIS AND PROTEIN FOLDING.			
CC	-1- SUBCELLULAR LOCATION: Mitochondrial.			
CC	-1- SIMILARITY: CONTAINS 1 J DOMAIN.			
CC	-1- SIMILARITY: CONTAINS 1 CR DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; Z28336; CAA82189.1; -			
DR	EMBL; Z46255; CAA86351.1; -			
DR	EMBL; D50617; BAA09222.1; -			
DR	PIR; S38898; S38898.			
DR	PIR; S48320; S48320.			
DR	HSP; P08622; IBOZ.			
DR	SGD; S0001878; MDJ1.			
DR	InterPro; IPR002939; DnaJ_C.			

Q9sjp6 arabidopsis  
P26758 chlamydia t  
P18151 chlamydia t  
P23603 chlamydia t  
P21354 chlamydia t  
Q61909 mus musculu  
Q06455 homo sapien  
Q38192 methanococ  
O15234 homo sapien  
Q9w04 mus musculu  
P70501 rattus norv  
Q00132 ictaluriid h

FU10\_ARATH  
OM6C\_CHLTR  
OM6D\_CHLTR  
OM6E\_CHLTR  
OM6L\_CHLTR  
MTG8\_MOUSE  
MTG8\_HUMAN  
TF2B\_METJA  
ML51\_HUMAN  
KIF9\_MOUSE  
RBM4\_RAT  
VG01\_HSV11



hypothetical protein YDR457w - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 22-Aug-1996 #sequence\_revision 06-Sep-1996 #text\_change 23-Mar-2001  
C:Accession: S69625  
R:Dietrich, F.S.  
submitted to the EMBL Data Library, August 1995  
A:Description: The sequence of S. cerevisiae cosmids 9410, 8035, 8166, and 9787.  
A:Reference number: S69554  
A:Accession: S69625  
A:Molecule type: DNA  
A:Residues: 1-3268 <DIE>  
A:Cross-references: EMBL:U33050; NID:g927726; PIDN:AAB64910.1; PID:g927738; MIPS:YDR457w  
C:Genetics:  
A:Gene: SGD:TOM1  
A:Cross-references: SGD:S0002865; MIPS:YDR457w  
A:Map position: 4R

Query Match 36.0%; Score 50; DB 2; Length 3268;  
Best Local Similarity 44.0%; Pred. No. 1.6e+02;  
Matches 11; Conservative 6; Mismatches 6; Indels 2; Gaps 1;

QY 2 KPNTNKEKAEKSHDTQTTOEICE 26  
|||:::| | |||||  
DB 226 KNNINETQTSIKV--TMTQELCE 248

## RESULT 15

T40036  
hypothetical protein SPBC27B12.12c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T40036  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lauber, J.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, December 1997  
A:Reference number: Z21900  
A:Accession: T40036  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-803 <WOO>  
A:Cross-references: EMBL:AL021766; PIDN:CAA16907.1; GSPDB:GN00067; SPDB:SPBC27B12.12c  
A:Experimental source: strain 972h.; cosmid c27B12  
C:Genetics:  
A:Gene: SPDB:SPBC27B12.12c  
A:Map position: 2

Query Match 35.3%; Score 49; DB 2; Length 803;  
Best Local Similarity 34.8%; Pred. No. 56;  
Matches 8; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 2 KPNTNKEKAEKSHDTQTTOEI 24  
||:|:| | |::|  
DB 150 KPQSNKKHGRVRVKHSPKSTLEV 172

Search completed: July 16, 2003, 18:10:00  
Job time : 5.06914 secs



C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

R;Accession: T25445  
R;Scheet, P.; Gattung, S.  
submitted to the EMBL Data Library, April 1997  
A:Description: The sequence of C. elegans cosmid B0261.  
A:Reference number: Z20036  
A:Accession: T25445  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-196 <SHA>  
A:Cross-references: EMBL:U97016; PIDN:AAB52352.1; GSPDB:GN00019; CESP:B0261.5  
A:Experimental source: strain Bristol N2; clone B0261  
C:Genetics:  
A:Gene: CESP:B0261.5  
A:Map position: 1  
A:Introns: 162/3

Query Match 36.0%; Score 50; DB 2; Length 196;  
Best Local Similarity 43.5%; Pred. No. 9.9;  
Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 2 KPTNTKSEKAERKSHDTQTTOEI 24  
Db 105 KASDRKKKANKKSPTQTQEV 127

RESULT 8  
JC5184  
glutamate synthase (GOGAT) (EC 1.4.1.1.) small chain - Thiobacillus ferrooxidans  
C:Species: thiobacillus ferrooxidans  
C>Date: 16-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 29-Sep-1999  
C:Accession: JC5184  
R:Deane, S.M.; Rawlings, D.E.  
Gene 177, 261-263, 1996  
A>Title: Cloning and sequencing of the gene for the Thiobacillus ferrooxidans ATCC33071  
A:Reference number: JC5184; MUID:97080532; PMID:8921877  
A:Accession: JC5184  
A>Status: preliminary; nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-465 <DEA>  
A:Cross-references: GB:U36427; NID:g1033069; PIDN:AAA79783.1; PID:g1033071  
A:Experimental source: strain ATCC33020  
C:Genetics:  
A:Gene: g1d  
C:Superfamily: glutamate synthase small chain  
C:Keywords: iron-sulfur protein; metalloprotein; oxidoreductase  
F:151-156,291-296/Region: glycine-rich  
F:427-437/Region: FAD-binding site predicted  
F:45,48,53,57,92,96,102,106/Binding site: iron-sulfur clusters (Cys) (covalent) #stat

Query Match 36.0%; Score 50; DB 2; Length 465;  
Best Local Similarity 50.0%; Pred. No. 23;  
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 6 NKSEKAERKSHDTQTTOEIC 25  
Db 73 NRLEEAATLSHQTTLPEIC 92

RESULT 9  
AD2545  
hypothetical protein alr7647 [Imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120  
C:Species: Nostoc sp.  
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AD2545  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AD2545  
A>Status: preliminary

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, K.; Ouchi, K.; Shiba, T.; Is  
Nucleic Acids Res. 28, 2311-2314, 2000  
A>Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349; PMID:10871362  
A:Accession: F86563  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-651 <STO>  
A:Cross-references: GB:BA000008; NID:g8978957; PIDN:BA98792.1; GSPDB:GN00142  
C:Genetics:  
A:Gene: CFJ0585

Query Match 36.7%; Score 51; DB 2; Length 651;  
Best Local Similarity 38.1%; Pred. No. 24;  
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 6 NKSEKAERKSHDTQTTOEICE 26  
Db 270 NOSQDKIQAHDRASQRACE 290

RESULT 6  
A72060  
hypothetical protein CP0163 [imported] - Chlamydophila pneumoniae (strains CWL029 and AR  
N:Alternate names: hypothetical protein CPn0585  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
C:Accession: A72060; B81607  
R:Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206506; PMID:10192388  
A:Accession: A72060  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-651 <ARN>  
A:Cross-references: GB:AE001643; GB:AE001363; NID:g4376876; PIDN:AAD18724.1; PID:g437687  
A:Experimental source: strain CWL029  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A>Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255; PMID:10684935  
A:Accession: B81607  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-651 <REA>  
A:Cross-references: GB:AE002177; GB:AE002161; NID:g7189090; PIDN:AAF38042.1; PID:g718909  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CPn0585; CP0163

Query Match 36.7%; Score 51; DB 2; Length 651;  
Best Local Similarity 38.1%; Pred. No. 24;  
Matches 8; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 6 NKSEKAERKSHDTQTTOEICE 26  
Db 270 NOSQDKIQAHDRASQRACE 290

RESULT 7  
T25445  
hypothetical protein B0261.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans

Cell 77, 249-259, 1994  
 A:Title: Mdjlp, a novel chaperone of the DnaJ family, is involved in mitochondrial biogenesis  
 A:Reference number: A53500; MUID:94221642; PMID:8168133  
 A:Accession: A53500  
 A:Molecule type: DNA  
 A:Residues: 1-511 <R02>  
 A:Cross-references: GB:Z28336; NID:g431909; PIDN:CAA82189.1; PID:g431910  
 R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, Y.  
 submitted to the EMBL Data Library, May 1995  
 A:Description: Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces cerevisiae*  
 A:Reference number: S56186  
 A:Accession: S56238  
 A:Molecule type: DNA  
 A:Residues: 1-511 <MUR>  
 A:Cross-references: EMBL:D50617; NID:g836685; PIDN:BAA09222.1; PID:g836738; MIPS:YFL016c  
 R:Murakami, Y.  
 submitted to the EMBL Data Library, December 1994  
 A:Reference number: S62230  
 A:Accession: S62297  
 A:Molecule type: DNA  
 A:Residues: 1-511 <MUW>  
 A:Cross-references: EMBL:D44596; NID:g1100783; PIDN:BAA08001.1; PID:g1100788  
 C:Genetics:  
 A:Gene: SGD:MDJ1  
 A:Cross-references: SGD:S0001878; MIPS:YFL016c  
 A:Map position: 6L  
 A:Genome: nuclear  
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
 C:Keywords: heat shock; membrane protein; mitochondrial inner membrane; mitochondrion; protein  
 F:1-55/Domain: transit peptide (mitochondrion) #status predicted <NP>  
 F:56-511/Product: heat shock protein MDJ1 #status predicted <MAY>  
 F:61-125/Domain: dnaJ amino-terminal homology <DNJ>  
 Query Match 40.3%; Score 56; DB 2; Length 511;  
 Best Local Similarity 50.0%; Pred. No. 3.6;  
 Matches 11; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
 Qy 3 PNTNKSKEAKRSHDQTOTQEI 24  
 I: | | | | | | | | | |  
 Db 90 PDINKEPDAKKFHDQNAVVEI 111  
 RESULT 3  
 S64758  
 SCD25 protein (version 2) - yeast (*Saccharomyces cerevisiae*)  
 A:Alternate names: protein L1309; protein YLL016w  
 C:Species: *Saccharomyces cerevisiae*  
 C:Date: 01-Aug-1995 #sequence.revision 24-May-1996 #text\_change 21-Jul-2000  
 C:Accession: S64758; S64764; S69390; S70559  
 R:Miosga, T.; Zimmermann, F.K.  
 submitted to the Protein Sequence Database, May 1996  
 A:Reference number: S64743  
 A:Accession: S64758  
 A:Molecule type: DNA  
 A:Residues: 1-1048 <MIO>  
 A:Cross-references: EMBL:D73121; NID:g1360186; PID:e245452; PID:g1360187; MIPS:YLL016w  
 A:Experimental source: strain S288C  
 R:Goffeau, A.; Purnelle, B.  
 submitted to the Protein Sequence Database, May 1996  
 A:Reference number: S64761  
 A:Accession: S64764  
 A:Molecule type: DNA  
 A:Residues: 1-1048 <GOF>  
 A:Cross-references: EMBL:D73121; NID:g1360186; PID:e245452; PID:g1360187; MIPS:YLL016w  
 A:Experimental source: strain S288C  
 R:Purnelle, B.; Goffeau, A.  
 submitted to the EMBL Data Library, April 1996  
 A:Description: The sequence of 32 kb on the left arm of yeast chromosome XII reveals 14 family and a new ABC transporter homologous to the human multidrug resistance protein.  
 A:Reference number: S69380  
 A:Accession: S69390  
 A:Molecule type: DNA  
 A:Residues: 1-1048 <PUR>

A:Cross-references: EMBL:X97560; NID:g1297003; PID:e238680; PID:g1297014  
 R:Miosga, T.; Zimmermann, F.K.  
 Yeast 12, 693-708, 1996  
 A:Title: Sequence analysis of the CEN12 region of *Saccharomyces cerevisiae* on a 43.7  
 e conductance regulator protein CFTR  
 A:Reference number: S70557; MUID:96405918; PMID:8810043  
 A:Accession: S70559  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1048 <MIW>  
 A:Cross-references: EMBL:X91488; NID:g1495203; PIDN:CAA62775.1; PID:g1495207  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 199  
 C:Genetics:  
 A:Gene: SGD:SCD25; SDC25  
 A:Cross-references: SGD:S0003939; MIPS:YLL016w  
 A:Map position: 12L  
 C:Superfamily: CDC25-type guanine nucleotide exchange activator homology  
 F:744-995/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>  
 Query Match 38.8%; Score 54; DB 2; Length 1048;  
 Best Local Similarity 45.0%; Pred. No. 14;  
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
 Qy 3 PNTNKSKEAKRSHDQTOTQ 22  
 I: | | | | | | | | | |  
 Db 1002 PNSNSNNKSQEKSRDQDTE 1021  
 RESULT 4  
 S14177  
 SCD25 protein (version 1) - yeast (*Saccharomyces cerevisiae*)  
 A:Alternate names: protein L1309; protein YLL016w  
 C:Species: *Saccharomyces cerevisiae*  
 C:Date: 02-Jun-1994 #sequence.revision 02-Jun-1994 #text\_change 06-Feb-1998  
 C:Accession: S14177; S12942; PS0040  
 R:Damak, F.; Boy-Marcotte, E.; Le-Roscouet, D.; Guilbaud, R.; Jacquet, M.  
 submitted to the EMBL Data Library, August 1989  
 A:Reference number: S14177  
 A:Accession: S14177  
 A:Molecule type: DNA  
 A:Residues: 1-1250 <DAM1>  
 A:Cross-references: EMBL:M26647  
 R:Damak, F.; Boy-Marcotte, E.; Le-Roscouet, D.; Guilbaud, R.; Jacquet, M.  
 Mol. Cell. Biol. 11, 203-212, 1991  
 A:Title: SCD25, a CDC25-like gene, which contains a RAS-activating domain and is a di  
 A:Reference number: S12942; MUID:91094833; PMID:1986220  
 A:Accession: S12942  
 A:Molecule type: DNA  
 A:Residues: 1-737, 'A', 738-970, 'I', 972-1250 <DAM2>  
 A:Cross-references: EMBL:M26647  
 A:Note: the authors translated the codon GAG for residue 538 as Asp, CTC for residue  
 R:Boy-Marcotte, E.; Damak, F.; Camonis, J.; Garreau, H.; Jacquet, M.  
 Gene 77, 21-30, 1989  
 A:Title: The C-terminal part of a gene partially homologous to CDC25 gene suppresses  
 A:Reference number: PS0041; MUID:89306677; PMID:2545538  
 A:Accession: PS0040  
 A:Molecule type: DNA  
 A:Residues: 668-679, 'NPVMTILMC', 689, 'N', 691-1250 <BOY>  
 A:Note: the authors translated the codon GCA for residue 747 as Asp and GAT for resid  
 C:Genetics:  
 A:Gene: SGD:SCD25; SDC25  
 A:Cross-references: SGD:S0003939; MIPS:YLL016w  
 A:Map position: 12L  
 C:Superfamily: CDC25-type guanine nucleotide exchange activator homology  
 F:946-1197/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>  
 Query Match 38.8%; Score 54; DB 2; Length 1250;  
 Best Local Similarity 45.0%; Pred. No. 17;  
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
 Qy 3 PNTNKSKEAKRSHDQTOTQ 22  
 I: | | | | | | | | | |  
 Db 1204 PNSNSNNKSQEKSRDQDTE 1223

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:50:14 ; Search time 3.06914 seconds  
(without alignments)  
814.396 Million cell updates/sec

Title: US-09-853-079-195  
Perfect score: 139  
Sequence: 1 GKPNTNKSEAKRSHDTQTQICE 26

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60.5	43.5	879	2 T22033	hypothetical prote
2	56	40.3	511	2 S38898	heat shock protein
3	54	38.8	1048	2 S4758	SCD25 protein (ver
4	54	38.8	1250	2 S4177	SCD25 protein (ver
5	51	36.7	651	2 F86563	hypothetical prote
6	51	36.7	651	2 A72060	hypothetical prote
7	50	36.0	196	2 T25445	hypothetical prote
8	50	36.0	465	2 J05184	glutamate synthase
9	50	36.0	471	2 A02545	hypothetical prote
10	50	36.0	579	2 D72092	conserved hypotet
11	50	36.0	579	2 C86532	CT082 hypothetical
12	50	36.0	923	2 G81253	probable oxidoredu
13	50	36.0	1409	2 T37188	presynaptic activi
14	50	36.0	3268	2 S89625	hypothetical prote
15	49	35.3	803	2 T40036	hypothetical prote
16	47.5	34.2	211	1 E69844	GRP pyrophosphokin
17	47	33.8	453	2 F71374	probable hemolysin
18	47	33.8	537	2 A86444	probable RNA helic
19	47	33.8	635	2 F71621	hypothetical prote
20	47	33.8	950	2 T09076	hypothetical prote
21	47	33.8	1097	2 T45622	hypothetical prote
22	46.5	33.5	65	2 G81292	hypothetical prote
23	46.5	33.5	1076	2 F96831	hypothetical prote
24	46	33.1	57	2 T10457	lipoprotein L - Ps
25	46	33.1	225	2 T28053	hypothetical prote
26	46	33.1	283	2 T23785	hypothetical prote
27	46	33.1	383	2 H90266	conserved hypotet
28	46	33.1	533	2 S63270	probable membrane
29	46	33.1	608	2 T25572	hypothetical prote

30 33.1 761 2 B69797 transcription requ  
31 46 788 2 C95046 helicase, probable  
32 46 831 2 T49758 related to cytochr  
33 45.5 631 2 T13115 protein gp29 - pha  
34 45.5 881 2 T33810 hypothetical prote  
35 45.5 890 2 T34243 hypothetical prote  
36 45.5 1230 2 T19899 hypothetical prote  
37 45 32.4 148 1 MCEG calmodulin - Eugle  
38 45 32.4 263 2 S01360 salivary glue prot  
39 45 32.4 310 2 T47695 hypothetical prote  
40 45 32.4 333 2 T19313 hypothetical prote  
41 45 32.4 542 2 T46464 hypothetical prote  
42 45 32.4 543 2 T02022 hypothetical prote  
43 45 32.4 749 2 A45294 Balbiani ring 2.1  
44 45 32.4 761 2 S66719 hypothetical prote  
45 45 32.4 798 2 S62791 probable lipoprote

#### ALIGNMENTS

##### RESULT 1

T22033

hypothetical protein F40F8.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T22033

R:MacDougall, R.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19505

A:Accession: T22033

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-879 <WIL>

A:Cross-references: EMBL:Z69302; PIDN:CAA93261.1; GSPDB:GNO0020; CESP:F40F8.5

A:Experimental source: clone F40F8

C:Genetics:

A:Gene: CESP:F40F8.5

A:Map position: 2

A:Introns: 63/1; 149/1; 173/1; 217/1; 262/1; 306/1; 347/1; 381/2; 422/2; 440/3; 461/3

Query Match 43.5%; Score 60.5; DB 2; Length 879;

Best Local Similarity 52.2%; Pred. No. 1.4;

Matches 12; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

Oy 5 TNKSEK-AERKSHDTQTQICE 26

||| || ||: ||: ||: ||

Db 3 TNKREKQLERECHEQKTSQAVCE 25

##### RESULT 2

S38898

heat shock protein MDJ1 precursor - yeast (Saccharomyces cerevisiae)

N:Alternate names: MDJ1 protein; protein YFL016C

C:Species: Saccharomyces cerevisiae

C:Date: 31-Dec-1993 #sequence\_revision 27-Jan-1995 #text\_change 20-Jun-2000

C:Accession: S38898; S43320; A53500; S56238; S62297

R:Rowley, N.K.; Prip-Buus, C.; Westermann, B.; Brown, C.M.; Schwarz, E.; Barrell, B.G

submitted to the EMBL Data Library, November 1993

A:Description: Mdj1p, a novel DnaJ homologue of Saccharomyces cerevisiae, plays a rol

A:Reference number: S38898

A:Accession: S38898

A:Molecule type: DNA

A:Residues: 1-511 <ROW>

A:Cross-references: EMBL:Z28336; NID:g431909; PIDN:CAA82189.1; PID:g431910

R:Churcher, C.

submitted to the EMBL Data Library, September 1994

A:Reference number: S48310

A:Accession: S48320

A:Molecule type: DNA

A:Residues: 1-511 <CHU>

A:Cross-references: EMBL:Z46255; NID:g559925; PIDN:CAA86351.1; PID:g559936; MIPS:YFL0

R:Rowley, N.; Prip-Buus, C.; Westermann, B.; Brown, C.; Schwarz, E.; Barrell, B.; Neu





; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C9  
; CURRENT APPLICATION NUMBER: US/09/737,178  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 87  
; LENGTH: 1132  
; TYPE: PRT  
; ORGANISM: Babesia  
US-09-737-178-87

Query Match 100.0%; Score 139; DB 10; Length 1132;  
Best Local Similarity 100.0%; Pred. NO. 1.4e-11;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GKPNTNKSEKAEKSHDTQTTOEICE 26  
Db 869 GKPNTNKSEKAEKSHDTQTTOEICE 894

Search completed: July 16, 2003, 18:08:38  
Job time : 4.25295 secs

## RESULT 11

US-09-737-178-85

; Sequence 85, Application US/09737178

; Patent No. US20010029295A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Sleath, Paul R.

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Homer, Mary

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION

; FILE REFERENCE: 210121.426C9

; CURRENT APPLICATION NUMBER: US/09/737,178

; CURRENT FILING DATE: 2000-12-13

; NUMBER OF SEQ ID NOS: 144

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 85

; LENGTH: 666

; TYPE: PRT

; ORGANISM: Babesia

US-09-737-178-85

Query Match 100.0%; Score 139; DB 10; Length 666;

Best Local Similarity 100.0%; Pred. No. 7.7e-12;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPTNKSEKAERKSHDTQTTOICE 26

| | | | | | | | | | | | | | | | | | | | | |

Db 403 GKPTNKSEKAERKSHDTQTTOICE 428

## RESULT 12

US-09-853-079-144

; Sequence 144, Application US/09853079

; Publication No. US20030109689A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Sleath, Paul R.

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Homer, Mary

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION

; FILE REFERENCE: 210121.426C11

; CURRENT APPLICATION NUMBER: US/09/853,079

; CURRENT FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 224

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 144

; LENGTH: 677

; TYPE: PRT

; ORGANISM: Babesia

US-09-853-079-144

Query Match 100.0%; Score 139; DB 9; Length 677;

Best Local Similarity 100.0%; Pred. No. 7.8e-12;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPTNKSEKAERKSHDTQTTOICE 26

| | | | | | | | | | | | | | | | | | | | | |

Db 175 GKPTNKSEKAERKSHDTQTTOICE 200

## RESULT 13

US-09-737-178-144

; Sequence 144, Application US/09737178

; Patent No. US20010029295A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C9  
; CURRENT APPLICATION NUMBER: US/09/737,178  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 144  
; LENGTH: 677  
; TYPE: PRT  
; ORGANISM: Babesia  
US-09-737-178-144

Query Match 100.0%; Score 139; DB 10; Length 677;

Best Local Similarity 100.0%; Pred. No. 7.8e-12;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPTNKSEKAERKSHDTQTTOICE 26

| | | | | | | | | | | | | | | | | | | | | |

Db 175 GKPTNKSEKAERKSHDTQTTOICE 200

## RESULT 14

US-09-853-079-87

; Sequence 87, Application US/09853079

; Publication No. US20030109689A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Sleath, Paul R.

; APPLICANT: McNeill, Patricia D.

; APPLICANT: Homer, Mary

; APPLICANT: Secrist, Heather

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION

; FILE REFERENCE: 210121.426C11

; CURRENT APPLICATION NUMBER: US/09/853,079

; CURRENT FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 224

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 87

; LENGTH: 1132

; TYPE: PRT

; ORGANISM: Babesia

US-09-853-079-87

Query Match 100.0%; Score 139; DB 9; Length 1132;

Best Local Similarity 100.0%; Pred. No. 1.4e-11;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPTNKSEKAERKSHDTQTTOICE 26

| | | | | | | | | | | | | | | | | | | | | |

Db 869 GKPTNKSEKAERKSHDTQTTOICE 894

## RESULT 15

US-09-737-178-87

; Sequence 87, Application US/09737178

; Patent No. US20010029295A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

```
; ORGANISM: Babesia microti
US-09-286-488-38

Query Match      100.0%; Score 139; DB 9; Length 445;
Best Local Similarity 100.0%; Pred. No. 4.9e-12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GKPNTNKSEKAERKSHDTQTTOEICE 26
   |||||
Db 182 GKPNTNKSEKAERKSHDTQTTOEICE 207

RESULT 7
US-09-853-079-38
; Sequence 38, Application US/09853079
; Publication No. US20030109689A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C11
; CURRENT APPLICATION NUMBER: US/09/853,079
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-853-079-38

Query Match      100.0%; Score 139; DB 9; Length 445;
Best Local Similarity 100.0%; Pred. No. 4.9e-12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GKPNTNKSEKAERKSHDTQTTOEICE 26
   |||||
Db 182 GKPNTNKSEKAERKSHDTQTTOEICE 207

RESULT 8
US-09-737-178-38
; Sequence 38, Application US/09737178
; Patent No. US20010029295A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-737-178-38

Query Match      100.0%; Score 139; DB 10; Length 445;
Best Local Similarity 100.0%; Pred. No. 4.9e-12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GKPNTNKSEKAERKSHDTQTTOEICE 26
   |||||
Db 403 GKPNTNKSEKAERKSHDTQTTOEICE 428

Query Match      100.0%; Score 139; DB 9; Length 666;
Best Local Similarity 100.0%; Pred. No. 7.7e-12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GKPNTNKSEKAERKSHDTQTTOEICE 26
   |||||
Db 403 GKPNTNKSEKAERKSHDTQTTOEICE 428
```

; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C11  
; CURRENT APPLICATION NUMBER: US/09/853.079  
; CURRENT FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 224  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 194  
; LENGTH: 50  
; TYPE: PRT  
; ORGANISM: B. microti  
US-09-853-079-194

Query Match 100.0%; Score 139; DB 9; Length 50;  
Best Local Similarity 100.0%; Pred. No. 4.3e-13;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPNTKSEKAERKSHDTQTTOICE 26  
|||||  
Db 1 GKPNTKSEKAERKSHDTQTTOICE 26

RESULT 3  
US-09-286-488=53  
; Sequence 53, Application US/09286488  
; Patent No. US20020169136A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C3  
; CURRENT APPLICATION NUMBER: US/09/286.488  
; CURRENT FILING DATE: 1999-04-05  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 53  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-286-488-53

Query Match 100.0%; Score 139; DB 9; Length 275;  
Best Local Similarity 100.0%; Pred. No. 2.9e-12;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPNTKSEKAERKSHDTQTTOICE 26  
|||||  
Db 218 GKPNTKSEKAERKSHDTQTTOICE 243

RESULT 4  
US-09-853-079-53  
; Sequence 53, Application US/09853079  
; Publication No. US20030109689A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C11

; CURRENT APPLICATION NUMBER: US/09/853.079  
; CURRENT FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 224  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 53  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-853-079-53

Query Match 100.0%; Score 139; DB 9; Length 275;  
Best Local Similarity 100.0%; Pred. No. 2.9e-12;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPNTKSEKAERKSHDTQTTOICE 26  
|||||  
Db 218 GKPNTKSEKAERKSHDTQTTOICE 243

RESULT 5  
US-09-737-178-53  
; Sequence 53, Application US/09737178  
; Patent No. US20010029295A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C9  
; CURRENT APPLICATION NUMBER: US/09/737.178  
; CURRENT FILING DATE: 2000-12-13  
; NUMBER OF SEQ ID NOS: 144  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 53  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-737-178-53

Query Match 100.0%; Score 139; DB 10; Length 275;  
Best Local Similarity 100.0%; Pred. No. 2.9e-12;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPNTKSEKAERKSHDTQTTOICE 26  
|||||  
Db 218 GKPNTKSEKAERKSHDTQTTOICE 243

RESULT 6  
US-09-286-488-38  
; Sequence 38, Application US/09286488  
; Patent No. US20020169136A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C3  
; CURRENT APPLICATION NUMBER: US/09/286.488  
; CURRENT FILING DATE: 1999-04-05  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 38  
; LENGTH: 445  
; TYPE: PRT

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:47:39 ; Search time 4.25295 Seconds  
(without alignments)  
726.026 Million cell updates/sec

Title: US-09-853-079-195

Perfect score: 139

Sequence: 1 GKPNTNKEAERKSHDTQTQICE 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	139	100.0	26	9	US-09-853-079-195
2	139	100.0	50	9	US-09-853-079-194
3	139	100.0	275	9	US-09-286-488-53
4	139	100.0	275	9	US-09-853-079-53
5	139	100.0	275	10	US-09-737-178-53
6	139	100.0	445	9	US-09-286-488-38
7	139	100.0	445	9	US-09-853-079-38
8	139	100.0	445	10	US-09-737-178-38
9	139	100.0	481	9	US-09-853-079-211
10	139	100.0	666	9	US-09-853-079-85
11	139	100.0	666	10	US-09-737-178-85
12	139	100.0	677	9	US-09-853-079-144
13	139	100.0	677	10	US-09-737-178-144
14	139	100.0	1132	9	US-09-853-079-87
15	139	100.0	1132	10	US-09-737-178-87
16	115	82.7	25	9	US-09-853-079-201
17	78	56.1	25	9	US-09-853-079-197
18	65	46.8	25	9	US-09-853-079-192
19	63	45.3	32	9	US-09-286-488-39

```

20 63 45.3 32 9 US-09-853-079-39 Sequence 39, Appl
21 63 45.3 32 10 US-09-737-178-39 Sequence 39, Appl
22 54 38.8 600 9 US-10-106-698-5866 Sequence 5866, Ap
23 51 36.7 25 9 US-09-853-079-202 Sequence 202, App
24 49 35.3 25 9 US-09-853-079-200 Sequence 200, App
25 49 35.3 121 9 US-10-078-090-119 Sequence 119, App
26 49 35.3 278 10 US-09-925-300-1323 Sequence 1323, Ap
27 47 33.8 178 9 US-10-174-590-542 Sequence 542, App
28 47 33.8 178 9 US-10-176-758-542 Sequence 542, App
29 47 33.8 178 9 US-10-175-737-542 Sequence 542, App
30 47 33.8 178 9 US-10-173-706-542 Sequence 542, App
31 47 33.8 178 9 US-10-175-738-542 Sequence 542, App
32 47 33.8 178 9 US-10-175-752-542 Sequence 542, App
33 47 33.8 178 9 US-10-176-482-542 Sequence 542, App
34 47 33.8 178 9 US-10-176-757-542 Sequence 542, App
35 47 33.8 178 9 US-10-176-913-542 Sequence 542, App
36 47 33.8 178 9 US-10-180-552-542 Sequence 542, App
37 47 33.8 178 9 US-10-180-557-542 Sequence 542, App
38 47 33.8 178 9 US-10-173-700-542 Sequence 542, App
39 47 33.8 178 9 US-10-174-572-542 Sequence 542, App
40 47 33.8 178 9 US-10-174-579-542 Sequence 542, App
41 47 33.8 178 9 US-10-174-582-542 Sequence 542, App
42 47 33.8 178 9 US-10-174-588-542 Sequence 542, App
43 47 33.8 178 9 US-10-175-739-542 Sequence 542, App
44 47 33.8 178 9 US-10-175-740-542 Sequence 542, App
45 47 33.8 178 9 US-10-175-743-542 Sequence 542, App

```

#### ALIGNMENTS

##### RESULT 1

```

US-09-853-079-195
; Sequence 195, Application US/09853079
; Publication No. US20030109689A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C11
; CURRENT APPLICATION NUMBER: US/09/853,079
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 195
; LENGTH: 26
; TYPE: PRT
; ORGANISM: B. microti
US-09-853-079-195

```

```

Query Match 100.0%; Score 139; DB 9; Length 26;
Best Local Similarity 100.0%; Pred. No. 2.1e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 GKPNTNKEAERKSHDTQTQICE 26
    |
Db 1 GKPNTNKEAERKSHDTQTQICE 26
    |

```

##### RESULT 2

```

US-09-853-079-194
; Sequence 194, Application US/09853079
; Publication No. US20030109689A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.

```



; LOCATION: (11)...(11)  
; OTHER INFORMATION: Xaa = Glutamic Acid or Glycine  
; NAME/KEY: VARIANT  
; LOCATION: (12)...(12)  
; OTHER INFORMATION: Xaa = Lysine or Asparagine  
; NAME/KEY: VARIANT  
; LOCATION: (14)...(14)  
; OTHER INFORMATION: Xaa = Glutamic Acid or Glycine  
; NAME/KEY: VARIANT  
; LOCATION: (15)...(15)  
; OTHER INFORMATION: Xaa = Isoleucine or Arginine  
; NAME/KEY: VARIANT  
; LOCATION: (18)...(18)  
; OTHER INFORMATION: Xaa = Histidine or Tyrosine  
; NAME/KEY: VARIANT  
; LOCATION: (23)...(23)  
; OTHER INFORMATION: Xaa = Threonine or Proline  
; NAME/KEY: VARIANT  
; LOCATION: (26)...(26)  
; OTHER INFORMATION: Xaa = Isoleucine or Threonine  
; NAME/KEY: VARIANT  
; LOCATION: (27)...(27)  
; OTHER INFORMATION: Xaa = Cysteine or Serine  
; NAME/KEY: VARIANT  
; LOCATION: (28)...(28)  
; OTHER INFORMATION: Xaa = Aspartic Acid or Glutamic Acid  
; NAME/KEY: VARIANT  
; LOCATION: (29)...(29)  
; OTHER INFORMATION: Xaa = Glutamic Acid or Alanine  
; NAME/KEY: VARIANT  
; LOCATION: (30)...(30)  
; OTHER INFORMATION: Xaa = Cysteine or Histidine  
US-09-528-784A-39

Query Match 45.3%; Score 63; DB 4; Length 32;  
Best Local Similarity 63.6%; Pred. No. 0.0046;  
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 2 KPNTNKSEKAERKSHDTQTTOE 23  
DB 4 KXNKNKXXAXXKSDTQTTOE 25

RESULT 14  
US-09-356-952-5  
; Sequence 5, Application US/09356952  
; Patent No. 6117663  
; GENERAL INFORMATION:  
; APPLICANT: Boriack-Sjodin, Ann  
; APPLICANT: Margarit, S. M.  
; APPLICANT: Bor-Sogi, Dafna  
; APPLICANT: Cole, Philip  
; APPLICANT: Kuriyan, John  
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE  
; FILE REFERENCE: 600-1-228N  
; CURRENT APPLICATION NUMBER: US/09/356,952  
; CURRENT FILING DATE: 1999-07-19  
; EARLIER APPLICATION NUMBER: 60/093,631  
; EARLIER FILING DATE: 1998-07-21  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1048  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-09-356-952-5

Query Match 38.8%; Score 54; DB 3; Length 1048;  
Best Local Similarity 45.0%; Pred. No. 5.9;  
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
QY 3 PNTNKSEKAERKSHDTQTTO 22

DB 1002 PNSNSNKNKSOEKSRRDQDTE 1021  
RESULT 15  
US-08-841-349-7  
; Sequence 7, Application US/08841349B  
; Patent No. 5955594  
; GENERAL INFORMATION:  
; APPLICANT: MISHRA, LOPA  
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.  
; FILE REFERENCE: XX/PO44700SU  
; CURRENT APPLICATION NUMBER: US/08/841,349B  
; CURRENT FILING DATE: 1997-04-30  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 220  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-08-841-349-7  
Query Match 33.8%; Score 47; DB 2; Length 220;  
Best Local Similarity 37.5%; Pred. No. 11;  
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;  
QY 3 PNTNKSEKAERKSHDTQTTOEICE 26  
DB 158 PNTKVSEEAESQOWDTSKGDQVSQ 181  
Search completed: July 16, 2003, 17:51:02  
Job time : 3.36762 secs



APPLICANT: Sleath, Paul R.  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
 TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
 NUMBER OF SEQUENCES: 49  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SEED AND BERRY  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/723,142A  
 FILING DATE: 01-OCT-1996  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Maki, David J.  
 REGISTRATION NUMBER: 31,392  
 REFERENCE/DOCKET NUMBER: 210121.426  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 39:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 32 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 3  
 OTHER INFORMATION: /note= "Residue can be either Gly  
 OTHER INFORMATION: or Asp"  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 5  
 OTHER INFORMATION: /note= "Residue can be either Pro  
 OTHER INFORMATION: or Ile"  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 7  
 OTHER INFORMATION: /note= "Residue can be either Lys  
 OTHER INFORMATION: or Thr"  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 11  
 OTHER INFORMATION: /note= "Residue can be either Glu  
 OTHER INFORMATION: or Gly"  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 12  
 OTHER INFORMATION: /note= "Residue can be either Lys  
 OTHER INFORMATION: or Asn"  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 14  
 OTHER INFORMATION: /note= "Residue can be either Glu  
 OTHER INFORMATION: or Gly"  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 15  
 OTHER INFORMATION: /note= "Residue can be either Ile  
 OTHER INFORMATION: or Arg"  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 18  
 OTHER INFORMATION: /note= "Residue can be either His  
 OTHER INFORMATION: or Trp"

```

;
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "Residue can be either Thr
; OTHER INFORMATION: Or Pro"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: Or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: Or Ser"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 28
; OTHER INFORMATION: /note= "Residue can be either Asp
; OTHER INFORMATION: Or Glu"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 29
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: Or Ala"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 30
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: Or His"
;
; US-08-723-142A-39
;
; Query Match 45.3%; Score 63; DB 4; Length 32;
; Best Local Similarity 63.6%; Pred. No. 0.0046;
; Matches 14; Conservative 0; Mismatches 8; Indels
;
; QY 2 KPNTNKSEKAERKSHDTQTQEE 23
; | | | | | | | | | | | | |
;
; Db 4 KXNKNKXXXXXXKSDTQTQXE 25
;
;
; RESULT 13
; US-09-528-784A-39
; Sequence 39, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; OF INFECTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
;
; SEQ ID NO 39
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Babesia microti
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3)...(3)
; OTHER INFORMATION: Xaa = Glycine or Aspartic Acid
; NAME/KEY: VARIANT
; LOCATION: (5)...(5)
; OTHER INFORMATION: Xaa = Proline or Isoleucine
; NAME/KEY: VARIANT
; LOCATION: (7)...(7)
; OTHER INFORMATION: Xaa = Lysine or Threonine
; NAME/KEY: VARIANT

```

```

; NAME/KEY: Modified-site
; LOCATION: 29
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: or Ala"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 30
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: or His"
US-08-845-258-39

```

```

Query Match 45.3%; Score 63; DB 4; Length 32;
Best Local Similarity 63.6%; Pred. No. 0.0046;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 2 KPNTNKSEKAERKSHDTQTQOE 23
   | | | | | | | | | | | | | | | |
Db 4 KXNKNKXXXXXXKSDTQTQOE 25

```

```

RESULT 11
US-08-990-571-39
; Sequence 39, Application US/08990571
; Patent No. 6214971
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,571
; FILING DATE: 11-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "Residue can be either Gly
; OTHER INFORMATION: or Asp"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5
; OTHER INFORMATION: /note= "Residue can be either Pro
; OTHER INFORMATION: or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 7
; OTHER INFORMATION: /note= "Residue can be either Lys
; OTHER INFORMATION: or Thr"
; FEATURE:

```

```

; NAME/KEY: Modified-site
; LOCATION: 11
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: or Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 12
; OTHER INFORMATION: /note= "Residue can be either Lys
; OTHER INFORMATION: or Asn"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: or Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: or Arg"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 18
; OTHER INFORMATION: /note= "Residue can be either His
; OTHER INFORMATION: or Tyr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "Residue can be either Thr
; OTHER INFORMATION: or Pro"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: or Ser"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 28
; OTHER INFORMATION: /note= "Residue can be either Asp
; OTHER INFORMATION: or Glu"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 29
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: or Ala"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 30
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: or His"
US-08-990-571-39

```

```

Query Match 45.3%; Score 63; DB 4; Length 32;
Best Local Similarity 63.6%; Pred. No. 0.0046;
Matches 14; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 2 KPNTNKSEKAERKSHDTQTQOE 23
   | | | | | | | | | | | | | | | |
Db 4 KXNKNKXXXXXXKSDTQTQOE 25

```

```

RESULT 12
US-08-723-142A-39
; Sequence 39, Application US/08723142A
; Patent No. 6306396
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond

```

```
; ORGANISM: Babesia
US-09-528-784A-85
Query Match      100.0%; Score 139; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 4.5e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPNTKSEKAERKSHDTQTQTEICE 26
    |||||
Db 403 GKPNTKSEKAERKSHDTQTQTEICE 428

RESULT 9
US-09-528-784A-87
; Sequence 87, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528.784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 87
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Babesia
US-09-528-784A-87

Query Match      100.0%; Score 139; DB 4; Length 1132;
Best Local Similarity 100.0%; Pred. No. 8.3e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPNTKSEKAERKSHDTQTQTEICE 26
    |||||
Db 869 GKPNTKSEKAERKSHDTQTQTEICE 894

RESULT 10
US-08-845-258-39
; Sequence 39, Application US/08845258
; Patent No. 6183976
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845.258
; FILING DATE: 24-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 3
; OTHER INFORMATION: /note= "Residue can be either Gly
; OTHER INFORMATION: Or Asp"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 5
; OTHER INFORMATION: /note= "Residue can be either Pro
; OTHER INFORMATION: Or Ile"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 7
; OTHER INFORMATION: /note= "Residue can be either Lys
; OTHER INFORMATION: Or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 11
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: Or Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 12
; OTHER INFORMATION: /note= "Residue can be either Lys
; OTHER INFORMATION: Or Asn"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 14
; OTHER INFORMATION: /note= "Residue can be either Glu
; OTHER INFORMATION: Or Gly"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 15
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: Or Arg"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 18
; OTHER INFORMATION: /note= "Residue can be either His
; OTHER INFORMATION: Or Tyr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 23
; OTHER INFORMATION: /note= "Residue can be either Thr
; OTHER INFORMATION: Or Pro"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 26
; OTHER INFORMATION: /note= "Residue can be either Ile
; OTHER INFORMATION: Or Thr"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 27
; OTHER INFORMATION: /note= "Residue can be either Cys
; OTHER INFORMATION: Or Ser"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 28
; OTHER INFORMATION: /note= "Residue can be either Asp
; OTHER INFORMATION: Or Glu"
; FEATURE:
```

```
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,571
; FILING DATE: 11-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-990-571-38
;
; Query Match 100.0%; Score 139; DB 4; Length 445;
; Best Local Similarity 100.0%; Pred. No. 2.8e-13;
; Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 GKPNTNKSEKAERKSHDTQTTOEICE 26
; Db 182 GKPNTNKSEKAERKSHDTQTTOEICE 207
;
; RESULT 6
; US-08-723-142A-38
; Sequence 38, Application US/08723142A
; Patent No. 6306396
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF B. MICROTI INFECTION
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/723,142A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
;
; Query Match 100.0%; Score 139; DB 4; Length 445;
; Best Local Similarity 100.0%; Pred. No. 2.8e-13;
; Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 GKPNTNKSEKAERKSHDTQTTOEICE 26
; Db 182 GKPNTNKSEKAERKSHDTQTTOEICE 207
;
; RESULT 7
; US-09-528-784A-38
; Sequence 38, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 445
; TYPE: PRT
; ORGANISM: Babesia microti
; US-09-528-784A-38
;
; Query Match 100.0%; Score 139; DB 4; Length 445;
; Best Local Similarity 100.0%; Pred. No. 2.8e-13;
; Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 GKPNTNKSEKAERKSHDTQTTOEICE 26
; Db 182 GKPNTNKSEKAERKSHDTQTTOEICE 207
;
; RESULT 8
; US-09-528-784A-85
; Sequence 85, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 85
; LENGTH: 666
; TYPE: PRT
```

```
Oy 1 GKPNKSEKAEKSHDTQTTOEICE 26
Db 218 GKPNKSEKAEKSHDTQTTOEICE 243

RESULT 2
US-09-528-784A-53
; Sequence 53, Application US/08990571
; Patent No. 6214971
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,571
; FILING DATE: 11-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Babesia microti
US-09-990-571-53

Query Match 100.0%; Score 139; DB 4; Length 275;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GKPNKSEKAEKSHDTQTTOEICE 26
Db 218 GKPNKSEKAEKSHDTQTTOEICE 243

RESULT 3
US-09-528-784A-53
; Sequence 53, Application US/09528784A
; Patent No. 6451115
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 53
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-528-784A-53

Query Match 100.0%; Score 139; DB 4; Length 275;
Best Local Similarity 100.0%; Pred. No. 1.6e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GKPNKSEKAEKSHDTQTTOEICE 26
Db 218 GKPNKSEKAEKSHDTQTTOEICE 243

RESULT 4
US-08-845-258-38
; Sequence 38, Application US/08845258
; Patent No. 6183976
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,258
; FILING DATE: 24-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 445 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-845-258-38

Query Match 100.0%; Score 139; DB 4; Length 445;
Best Local Similarity 100.0%; Pred. No. 2.8e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GKPNKSEKAEKSHDTQTTOEICE 26
Db 182 GKPNKSEKAEKSHDTQTTOEICE 207

RESULT 5
US-08-990-571-38
; Sequence 38, Application US/08990571
; Patent No. 6214971
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:42:54 ; Search time 2.36762 seconds  
(without alignments)  
323.107 Million cell updates/sec

Title: US-09-853-079-195

Perfect score: 139  
Sequence: 1 GKPNTKSEKAEKSHDTQTQICE 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA:\*

- 1: /cgn2.6/ptodata/1/1aa/5A\_COMB.pep.\*
- 2: /cgn2.6/ptodata/1/1aa/5B\_COMB.pep.\*
- 3: /cgn2.6/ptodata/1/1aa/6A\_COMB.pep.\*
- 4: /cgn2.6/ptodata/1/1aa/6B\_COMB.pep.\*
- 5: /cgn2.6/ptodata/1/1aa/PT05\_COMB.pep.\*
- 6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	139	100.0	275	4	US-08-845-258-53
2	139	100.0	275	4	US-08-990-571-53
3	139	100.0	275	4	US-09-528-784A-53
4	139	100.0	445	4	US-08-845-258-38
5	139	100.0	445	4	US-08-990-571-38
6	139	100.0	445	4	US-08-723-142A-38
7	139	100.0	445	4	US-09-528-784A-38
8	139	100.0	666	4	US-09-528-784A-85
9	139	100.0	1132	4	US-09-528-784A-87
10	63	45.3	32	4	US-08-845-258-39
11	63	45.3	32	4	US-08-990-571-39
12	63	45.3	32	4	US-08-723-142A-39
13	63	45.3	32	4	US-09-528-784A-39
14	54	38.8	1048	3	US-09-356-952-5
15	47	33.8	220	2	US-08-841-349-7
16	47	33.8	2154	2	US-08-841-349-4
17	46	33.1	259	4	US-08-185-160-9
18	46	33.1	734	4	US-09-185-160-13
19	45.5	32.7	219	4	US-09-134-001C-5651
20	44	31.7	343	1	US-08-348-792-10
21	44	31.7	343	2	US-08-462-738-10
22	44	31.7	343	4	US-09-199-955-10
23	44	31.7	343	4	US-08-880-875-10
24	43	30.9	123	1	US-08-131-625B-17
25	43	30.9	123	5	PCT-US95-10904-19
26	43	30.9	534	2	US-08-691-814B-8
27	43	30.9	576	4	US-09-134-001C-3955

28	43	30.9	652	4	US-09-134-001C-3517	Sequence 3517, Ap
29	43	30.9	731	4	US-09-185-160-11	Sequence 11, Appl
30	43	30.9	752	1	US-08-244-189-2	Sequence 2, Appli
31	43	30.9	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
32	42	30.2	245	4	US-08-469-260A-37	Sequence 37, Appl
33	42	30.2	461	2	US-08-463-587A-26	Sequence 26, Appl
34	42	30.2	461	3	US-08-463-667A-4	Sequence 4, Appli
35	42	30.2	461	3	US-08-923-854-26	Sequence 26, Appl
36	42	30.2	461	5	PCT-US91-09133-27	Sequence 27, Appl
37	42	30.2	546	4	US-09-352-990-2	Sequence 2, Appli
38	42	30.2	588	2	US-08-459-346-12	Sequence 12, Appl
39	42	30.2	588	3	US-08-889-419-12	Sequence 12, Appl
40	42	30.2	588	4	US-08-402-542-12	Sequence 12, Appl
41	42	30.2	588	5	PCT-US93-07189-12	Sequence 12, Appl
42	42	30.2	1566	2	US-08-687-956A-23	Sequence 23, Appl
43	42	30.2	1848	4	US-08-296-791-6	Sequence 6, Appli
44	42	30.2	1848	5	PCT-US95-10661A-6	Sequence 6, Appli
45	41.5	29.9	123	2	US-08-799-464A-13	Sequence 13, Appl

#### ALIGNMENTS

RESULT 1  
US-08-845-258-53  
; Sequence 53, Application US/08845258  
; Patent No. 6183976  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/845,258  
; FILING DATE: 24-APR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.426C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 275 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Babesia Microti  
; US-08-845-258-53

Query Match: 100.0%; Score 139; DB 4; Length 275;  
Best Local Similarity 100.0%; Pred. No. 1.6e-13;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Search completed: July 16, 2003, 17:50:04  
Job time : 125.145 secs









```

Qy 211 SAALFRYKEDDELFFKKNANFGDDGYDFDYINTKK-ELVI--LASVLDGLDLMERLI 267
Db 370 SFCFFVYKYEHLARLHNEERCANVKDKIKNLEYVQGLGTIINA--NIVEYLV 427
Qy 268 ENFSDVNTDDIKKAFDECKSNAILKKILDNDEYKINFREMVNEV--TCANTFEAL 325
Db 428 --FEDEPLTTIGFRYTDKSOONELSHPNKI-----YKIFSDYIIEFDDDAKLTTGTG 479
Qy 326 NDLIISCEKKGKINRDVSISSYKLLISTITY 357
Db 480 EDITIYTC-----KHNPVLIRFSCSIEKYY 506

RESULT 7
Q9LH90 PRELIMINARY; PRT; 342 AA.
AC Q9LH90:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL: AP002057; BAB03182.1; -.
SQ SEQUENCE 342 AA; 31165 MW; B78F9602BCD914A7 CRC64;

Query Match 8.1%; Score 209; DB 10; Length 342;
Best Local Similarity 40.1%; Pred. No. 0.0081;
Matches 69; Conservative 24; Mismatches 57; Indels 22; Gaps 12;

Qy 323 EALNDLIISDCEKKGKINRDVSISSYKLLISTI-TYIVGAGVAVTVSVSATSNGTSGG 381
Db 111 OTIGQFVYSAAKSGNKID---ISS--LGIDGIDASAGAGDSTAGGSVAT--GTGSYG 163
Qy 382 AGSGTGSVATSLTNGGTSGGT---AGTTTSSGTAGGTSGT---TSSGASGK 434
Db 164 AGAG-GSSAGSDTAAG-GTATGTTAGGTAAGGTTAGG-TAGTACATAGGATAG 220
Qy 435 AGTGATAGTTSSGG--AGSDKAGTGT-SGTTTSSGTG-AGGAGGPGPSGHAS 482
Db 221 AGTAAGGTTADGATAGTATAGTAAAGGTTAAGGTTAAGAGAAAGAGAGAGAA 272

RESULT 8
Q9D2K8 PRELIMINARY; PRT; 637 AA.
AC Q9D2K8:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 0 day neonate head cDNA, RIKEN full-length enriched library,
DE clone:4833436C19, full insert sequence.
GN KRT2-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK019521; BAB31776.1; -.
DR MGI: MGI:96698; Krt2-1.
DR InterPro: IPR001664; IF.
DR InterPro: IPR002957; Keratin_I.
DR InterPro: IPR003054; Keratin_II.
DR InterPro: IPR000817; Prion.
DR Pfam: PF00038; filament; 1.
DR PRINTS: PR00341; PRION.
DR PRINTS: PR01248; TYPE1KERATIN.
DR PRINTS: PR01276; TYPE2KERATIN.
DR PROSITE: PS00226; IF; UNKNOWN_1.
SQ SEQUENCE 637 AA; 65627 MW; D6344FDD0468EA7 CRC64;

Query Match 8.1%; Score 207.5; DB 11; Length 637;
Best Local Similarity 21.5%; Pred. No. 0.019;
Matches 110; Conservative 74; Mismatches 175; Indels 153; Gaps 18;

Qy 85 GVSDFPIRQNLGAVGSV-----NEQNTVGMLEQPIKNELYSFSNEIYHTTSSQISN- 139
Db 136 GFSGGFGGFGGTGCGMGPVCPGGIQEVTINQSLLOPLNVEVDPQIQKVSQEREKSL 195
Qy 140 -----SPLIMSDAIVKHDNVILKKEGCEQIYNYEEFIEKLGARSEGNNMFQALIR 194
Db 196 NDXFASPIDKV--RFLEQQNOVLQTKWELLQOQVDT-----TTQTQNDLPFFENYIS 244
Qy 195 FRNASSEMYNAASYLSAALFRYKEDDELFFKKNANFGDDGYDFDYINTKKEL----- 249
Db 245 ILRRKVDLSLSDQSRMDSLEKNQDLYVEEYRTKYEDEINKRTNAENEFVTKKDVSAYM 304
Qy 250 --VLASVLDGLDLMERLIENFS-----DVNNTDDIKKAFD 284
Db 305 TKVELOKADAL---QQDIDFSSALYQMEMSQMOTQISETNVVLSMDNRSRLDGLGIS 360
Qy 285 ECKSN-AIILKKILDNDEYKINFREMVNEVTC-----NTKFE--ALNDLI----- 329
Db 361 EVKQYDSCICORSKAAEATFYQSKYEEL--QITAGKHGDSVRNRTKMEISELNRIQLRS 418
Qy 330 -----ISDCEKKGKIKINRDV----- 344
Db 419 EIDGCKKQIYQIQONINDAEGEKAALKDAQNKLNEIEDALSQCKEDLARLLRDFQELMN 478
Qy 345 -----ISSYKLLSTITYIVGAGVE-----AVTVSVSATS-NGTSGGAGS 384
Db 479 TKLALDMEIATYKKLLE-----GGEIRMSGECTPNVSVSVSTSHTSMSGSSRRGGS 530
Qy 385 GTGTSVSATSLTNGGTSGGTAGTTTSSGTAGGTGTTSSGTAGGASGAGTGTACTTTT 444

```



Db 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEFIEKLRA 180  
QY 181 RSEGNMFQALIRFRNASSEEMVNAASLSAALFRYKEFDDELFKKANDNFGRDDGYDF 240  
Db 181 RSEGNMFQALIRFRNASSEEMVNAASLSAALFRYKEFDDELFKKANDNFGRDDGYDF 240  
QY 241 DYINTKKELVILASVLDGLDLIMERLIENFSDVNNNTDDIKKAFDECKSNAILKKKILDN 300  
Db 241 DYINTKKELVILASVLDGLDLIMERLIENFSDVNNNTDDIKKAFDECKSNAILKKKILDN 300  
QY 301 DEDYKINFRMNEVTCANTKPEALNDLIISDCEKKGKINRDVIVSSYKLLSTIYIVG 360  
Db 301 DEDYKINFRMNEVTCANTKPEALNDLIISDCEKKGKINRDVIVSSYKLLSTIYIVG 360  
QY 361 AGVEAVTVSVATSNCTSGGAGSGTGTSVTSATSTLTGNGGTSGGTAGTTSSSGTEAGG 420  
Db 361 AGVEAVTVSVATSNCTSGGAGSGTGTSVTSATSTLTGNGGTSGGTAGTTSSSGTEAGG 420  
QY 421 TSCTTTSSGAAGKAGTGTAGTTSSGAGSDKAGTGTGTTSSGAGGAGSGGPGSGH 480  
Db 421 TSCTTTSSGAAGKAGTGTAGTTSSGAGSDKAGTGTGTTSSGAGGAGSGGPGSGH 480  
QY 481 ASNAKIPGIMTLFLFALLTFIVN 503  
Db 481 ASNAKIPGIMTLFLFALLTFIVN 503

RESULT 2  
Q9NIN9 ID Q9NIN9 PRELIMINARY; PRT; 492 AA.  
AC Q9NIN9  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE Seroreactive antigen BMN1-4 (Fragment).  
OS Babesia microti.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
OX NCBI\_TaxID=5868;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MNI;  
RX MEDLINE=20231818; PubMed=10768973;  
RA Lodes M.J., Houghton R.L., Bruinsma E.S., Mohamath R., Reynolds L.D.,  
RA Benson D.R., Krause P.J., Reed S.G., Persing D.H.;  
RA "Serological expression cloning of novel immunoreactive antigens of  
RT Babesia microti."  
RL Infect. Immun. 68:2783-2790(2000).  
DR EMBL; AF206246; AAF68238.1; -;  
FT NON\_TER  
FT SEQUENCE 492 AA; 55235 MW; A801E9CACC35881 CRC64;

Query Match 83.0%; Score 2130; DB 5; Length 492;  
Best Local Similarity 99.5%; Pred. No. 1.9e-93;  
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KRNEHTDMNGIHYIIDGSLASGEVTSNFRYISKEYEYEHTELAKHCKKCKVNDN 60  
Db 73 KRNEHTDMNGIHYIIDGSLASGEVTSNFRYISKEYEYEHTELAKHCKKCKVNDN 132  
QY 61 IEDNNLKIYAKQFQSVVTTADVAGVSDGFFIRGQNLGAVGSNEQPTVGMSELOFIKN 120  
Db 133 IEDNNLKIYAKQFQSVVTTADVAGVSDGFFIRGQNLGAVGSNEQPTVGMSELOFIKN 192  
QY 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEFIEKLRA 180  
Db 193 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEFIEKLRA 252  
QY 181 RSEGNMFQALIRFRNASSEEMVNAASLSAALFRYKEFDDELFKKANDNFGRDDGYDF 240  
Db 253 RSEGNMFQALIRFRNASSEEMVNAASLSAALFRYKEFDDELFKKANDNFGRDDGYDF 312  
QY 241 DYINTKKELVILASVLDGLDLIMERLIENFSDVNNNTDDIKKAFDECKSNAILKKKILDN 300

Db 313 DYINTKKELVILASVLDGLDLIMERLIENFSDVNNNTDDIKKAFDECKSNAILKKKILDN 372  
QY 301 DEDYKINFRMNEVTCANTKPEALNDLIISDCEKKGKINRDVIVSSYKLLSTIYIVG 360  
Db 373 DEDYKINFRMNEVTCANTKPEALNDLIISDCEKKGKINRDVIVSSYKLLSTIYIVG 432  
QY 361 AGVEAVTVSVATSNCTSGGAGSGTGTSVTSATSTLTGNGGTSGGTAGTTSSSGTEAG 419  
Db 433 AGVEAVTVSVATSNCTSGGAGSGTGTSVTSATSTLTGNGGTSGGTAGTTSSSGTWF 491

RESULT 3  
Q9NIP1 ID Q9NIP1 PRELIMINARY; PRT; 362 AA.  
AC Q9NIP1  
DT 01-OCT-2000 (TReMBLrel. 15, Created)  
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)  
DE Seroreactive antigen BMN1-3.  
OS Babesia microti.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
OX NCBI\_TaxID=5868;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MNI;  
RX MEDLINE=20231818; PubMed=10768973;  
RA Lodes M.J., Houghton R.L., Bruinsma E.S., Mohamath R., Reynolds L.D.,  
RA Benson D.R., Krause P.J., Reed S.G., Persing D.H.;  
RA "Serological expression cloning of novel immunoreactive antigens of  
RT Babesia microti."  
RL Infect. Immun. 68:2783-2790(2000).  
DR EMBL; AF206245; AAF68236.1; -;  
SQ SEQUENCE 362 AA; 39988 MW; 1B88DCFEF2BD5133 CRC64;  
Query Match 9.3%; Score 239.5; DB 5; Length 362;  
Best Local Similarity 29.3%; Pred. No. 0.00031;  
Matches 85; Conservative 50; Mismatches 96; Indels 59; Gaps 11;  
QY 107 PNTVGMSLQFTKNELYSFSNEYHTIYHTISSQISNSFLIMMSDAIVK-----HDNYI--- 156  
Db 99 PSEAGWSSERF-GYQLLPYSRRI--VIFNEVCLSIYIKHSVMILRDRVNDGDKDYIEBK 155  
QY 157 -----LKKEGEGCEQIYNYEFIEKLRCARSGNNMFQALIR-PRNASSEEMVNAASY 209  
Db 156 TEKNKLKKELEKC-----FPEQY-----SLMKKEELARIPDNAST----- 191  
QY 210 LSAALFRYKEFDDELFKKANDNFGRDDGYDFDIYNTKKELVILASVLDGLDLIMERLIEN 269  
Db 192 ISS---KYLLVDEISNKAYGTLEGPAADNFDHFRNIWKSIVLKDIFYCDLLLOHLYK 248  
QY 270 FSDVNNNTDDIKKAFDECKSNAILKKKILONDEYKINFRMNEVTCANTKPEALNDLI 329  
Db 249 FYDNTVNDIKKNFDESKALVLRDKITKKGDDYNTHPFEDMIKELNSAAEFNKIVDIM 308  
QY 330 ISDCEKKGKINRDVIVSSYKLLSTIYI-----VCAGVEAVTVSV 370  
Db 309 ISNI-----GDYDEYDSTASEKPFUSMITEITKITKVSNNVIIPGKALTIV 355

RESULT 4  
O04310 ID O04310 PRELIMINARY; PRT; 705 AA.  
AC O04310  
DT 01-JUL-1997 (TReMBLrel. 04, Created)  
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE Jasmonate inducible protein ISOLOG (Jasmonate inducible protein,  
DE myrosinase binding protein-like) (Putative jasmonate inducible  
DE protein).  
DE T02004.4.  
GN  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:37:48 ; Search time 122.145 Seconds  
(without alignments)  
848.513 Million cell updates/sec

Title: US-09-853-079-52

Perfect score: 2565

Sequence: 1 KRNEHTDMNGIHYHYIDGS.....AKIPGIMTLTLPALLTFIVN 503

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2565	100.0	503	5 Q9NIN8	Q9nln8 babesia mic
2	2130	83.0	492	5 Q9NIN9	Q9nin9 babesia mic
3	239.5	9.3	362	5 Q9NIP1	Q9nlp1 babesia mic
4	234.5	9.1	705	10 O04310	O04310 arabidopsis
5	216	8.4	680	10 P93658	P93658 brassica na
6	212	8.3	592	5 Q9NIN3	Q9nin3 babesia mic
7	209	8.1	342	10 Q9LH90	Q9lnh90 arabidopsis
8	207.5	8.1	637	11 Q9D2K8	Q9d2k8 mus musculus
9	200.5	7.8	1428	5 O44341	O44341 haliotis ru
10	199.5	7.8	138	5 Q964C4	Q964c4 encephalito
11	198.5	7.7	614	10 Q9LH97	Q9lnh97 arabidopsis
12	198	7.7	1079	16 O35557	O35557 mycobacteri
13	197.5	7.7	561	4 Q14664	Q14664 homo sapien
14	197	7.7	125	5 Q964C1	Q964c1 encephalito
15	196.5	7.7	1218	5 Q9W4X0	Q9w4x0 drosophila
16	195	7.6	707	11 Q61869	Q61869 mus musculus

```

17 194.5 7.6 561 11 Q9CXH6
18 191.5 7.5 117 5 Q964C2
19 191 7.4 461 5 Q9NIN0
20 190 7.4 426 16 Q8Y293
21 188.5 7.3 251 3 Q9UVE7
22 187 7.3 1713 3 Q8TGE1
23 186 7.3 121 5 Q964C5
24 186 7.3 1489 16 O53559
25 185 7.2 1217 16 Q8VIY9
26 184 7.2 967 3 Q08294
27 183.5 7.2 619 12 Q9IPQ9
28 183.5 7.2 1381 16 O53552
29 183.5 7.2 1384 16 Q8VIZ1
30 182.5 7.1 1536 2 Q48031
31 182 7.1 233 5 Q9U0C2
32 182 7.1 1357 5 Q9W4M4
33 181.5 7.1 786 6 O18740
34 181 7.1 853 2 Q8VQ20
35 181 7.1 857 2 O85783
36 181 7.1 1001 3 Q05164
37 181 7.1 1306 16 O53775
38 180 7.0 215 5 Q9L8W1
39 180 7.0 3178 5 Q969D4
40 179 7.0 292 5 Q9NIM9
41 179 7.0 635 16 Q8VK71
42 179 7.0 775 16 Q8VJ15
43 179 7.0 1715 16 Q8VIZ0
44 178.5 7.0 487 2 Q92B39
45 178 6.9 592 16 Q9PF60

```

#### ALIGNMENTS

#### RESULT 1

```

Q9NIN8 PRELIMINARY; PRT; 503 AA.
ID Q9NIN8
AC Q9NIN8;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Seroreactive antigen MN-10 (Fragment).
OS Babesia microti.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.
OX NCBI_TaxID=5868;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MN1;
RX MEDLINE=20231818; PubMed=10768973;
RA Lodes M.J., Houghton R.L., Bruinsma E.S., Mohamath R., Reynolds L.D.,
RA Benson D.R., Krause P.J., Reed S.G., Persing D.H.;
RT "Serological expression cloning of novel immunoreactive antigens of
RT Babesia microti."
RL Infect. Immun. 68:2783-2790(2000).
DR EMBL; AF206247; AAF68239.1; -.
DR HSSP; P00778; IGBJ.
FT NON_TER 1
SQ SEQUENCE 503 AA; 53831 MW; 0098D5DBF1BEF5AA CRC64;

Query Match 100.0%; Score 2565; DB 5; Length 503;
Best Local Similarity 100.0%; Pred. No. 5,2e-114;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KRNEHTDMNGIHYHYIDGSLASGEVTSNFRYISKEYEYHTELAKHCKKCVNVDN 60
Db 1 KRNEHTDMNGIHYHYIDGSLASGEVTSNFRYISKEYEYHTELAKHCKKCVNVDN 60
QY 61 IEDNKLKIYAKQFKSVVITPADVAGVSDGFFTRGQNLGAVGSVNQPNVTGMSLQFIKN 120
Db 61 IEDNKLKIYAKQFKSVVITPADVAGVSDGFFTRGQNLGAVGSVNQPNVTGMSLQFIKN 120
QY 121 ELYSFSNEIYHTYISSQISNSFLIMSDAIVKHNDYTLKKEGCGCEQIYNYEEFIEKLRA 180

```









DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Hypothetical PE-PGRS family protein RV3508 precursor.
GN	RV3508 OR MV023.15.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX	NCLB_TaxID=1773;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=H37Rv;
RX	MEDLINE=98295987; PubMed=9634230;
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekait F.,
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA	Oliver S., Osborne J.J., Quail M.A., Rajandream M.A., Rogers J.,
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.:
RT	"Deciphering the biology of Mycobacterium tuberculosis from the
RL	complete genome sequence.";
RL	Nature 393:537-544(1998).
CC	-!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC	SUBFAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or
CC	send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AL022022; CAAL17745.1; -
DR	TubercuList; RV3508; -
DR	InterPro; IPR000084; PE_region.
DR	Pfam; PF00934; PE; 1.
DR	ProDom; PD001223; PE_region; 1.
KW	Hypothetical protein; Repeat; Signal; Complete proteome.
FT	SIGNAL 1 30 POTENTIAL.
FT	CHAIN 31 1901 HYPOTHETICAL PE-PGRS* FAMILY PROTEIN
FT	RV3508.
SQ	SEQUENCE 1901 AA; 147627 MW; C7B1923D5D0146CD CRC64;
Query Match	7.2%; Score 185; DB 1; Length 1901;
Best Local Similarity	37.88; Pred.No. 0.02;
Matches	54; Conservative 12; Mismatches 55; Indels 22; Gaps 6;
QY	360 GAGVEAVTVSVSATSNQTE-SGGA----SGTGTSVSATSLTLNGGT-----ESGGTAGT 410      :   :         :   :   :   :   :   :   :   :
Db	513 GAGSAGTDATACATGTGTCFGSGAGGAGGAGNTGVCGTNGSGQGTGAGGAGGAGV 572      :   :         :   :   :   :   :   :   :   :
QY	411 TTSSGTAGTSTTTSSSA--ASGKAGTACTATTTSSEGAAGSKAGTGTGTTTTSSG-- 466 :          :   :   :   :   :   :   :   :   :
Db	573 GADNPVTIGTGTGTGGKAGGAGGQGGSAGGCTNGSGAG---GTGQGAGGAGGAGA 628            :   :   :   :   :   :   :   :   :
QY	467 -----TCAGAGSGGPSGHASNA 484
Db	629 GADNPVTIGGAGGTGTGGGAAGA 651
RESULT 13	
QY34_MYC TU	ID YQ34_MYC TU STANDARD; PRP; 778 AA.
ID	YQ34_MYC TU AC P71933;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Hypothetical PE-PGRS family protein RV2634c.
GN	RV2634C OR MT2172 OR MTCI441.04C.
OS	Mycobacterium tuberculosis.
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

```
or send an email to license@isb-sib.ch).
-----
CC EMBL; L00193; AAA39391.1; -.
DR EMBL; M10081; AAA39391.1; JOINED.
DR EMBL; V00830; CAA24214.1; -.
DR PIR; A02940; KRMS1.
DR PIR; S07330; S07330.
DR MGD; MGI:96685; Krt1-10.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament_1.
DR PRINTS; PR01248; TYPE1KERATIN.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin.
FT INIT_MET 0
FT DOMAIN 1 142 HEAD.
FT DOMAIN 143 453 ROD.
FT DOMAIN 454 569 TAIL.
FT DOMAIN 143 178 COIL 1A.
FT DOMAIN 179 199 LINKER 1.
FT DOMAIN 200 291 COIL 1B.
FT DOMAIN 292 314 LINKER 12.
FT DOMAIN 315 453 COIL 2.
FT SITE 395 395 STUTTER.
FT DOMAIN 452 564 GLY/SER-RICH.
FT CONFLICT 5 5 S -> C (IN REF. 2).
FT CONFLICT 24 24 S -> F (IN REF. 2).
FT CONFLICT 28 28 S -> F (IN REF. 2).
FT CONFLICT 38 38 Y -> L (IN REF. 2).
FT CONFLICT 41 41 E -> G (IN REF. 2).
FT CONFLICT 104 105 AG -> GS (IN REF. 2).
FT CONFLICT 110 110 MISSING (IN REF. 2).
FT CONFLICT 121 122 SY -> GC (IN REF. 2).
FT CONFLICT 137 137 S -> G (IN REF. 2).
FT CONFLICT 148 148 Q -> R (IN REF. 2).
FT CONFLICT 178 187 WYKKGNSO -> VVREAROLKP (IN REF. 2).
FT CONFLICT 263 268 KSDLEM -> QSVLEL (IN REF. 2).
FT CONFLICT 284 284 H -> L (IN REF. 2).
FT CONFLICT 353 353 E -> A (IN REF. 2).
FT CONFLICT 394 399 EGRYCV -> VESLLR (IN REF. 2).
FT CONFLICT 508 514 GGSFGGS -> CGGRGG (IN REF. 2).
FT CONFLICT 523 523 S -> G (IN REF. 2).
FT CONFLICT 531 531 H -> R (IN REF. 2).
FT CONFLICT 534 534 S -> G (IN REF. 2).
FT CONFLICT 543 543 S -> G (IN REF. 2).
FT CONFLICT 547 548 GQ -> RR (IN REF. 2).
FT CONFLICT 555 556 KS -> SGT (IN REF. 2).
SQ SEQUENCE 569 AA; 57711 MW; EEC59D4D8FEB484D CRC64;

Query Match 7.5%; Score 192; DB 1; Length 569;
Best Local Similarity 22.0%; Pred. No. 0.002;
Matches 111; Conservative 68; Mismatches 192; Indels 134; Gaps 19;

QY 19 GSLLA-SGEVTS---NFR---YISKEVEYHTELAKHCKKCVNVDNIEDNNLKIVAK 71
DB 136 GSLSGNGRVTMQNLNDRLASYMDKVRALAESNVELEGKIKWEYKKGNSQRPRDYK 195
QY 72 QFKS-----VVTTPADVAGV-----SDGFFIRGNLGAAGSVNEQPTVGMSL 114
DB 196 YKKTIEDLKGQILITLTDNANVLQIDNARLAADDFRLKYEN-----EVTLRQSV 245
QY 115 EQFTKNELYSFSEIYHTISSQNSFLIMMSDAIVKHNDNVLKKEGCGQIYNYE--- 171
DB 246 EADI-NGLRRVLDEL-----TLKSDLEMQIESLNEELAYLKKHHEEMRDLQNVSTGD 298
QY 172 -----EFTKLRGARSEGNMFOEALIFRNASSEMVAASLYSAALFRYKEF 220
DB 299 VNVEMNAAPGVLDLTQLLNMRNQ-----YEQLAEKNRKDAEWFNOKS----- 341
QY 221 DDELFKKANDNFGDDGDFDYINTKKELVTLASVLDGLDLIMERLTENFSDVNNDDIK 280
DB 342 -KELTTEIDSNIEQWSSH-----KSEITELRRTVOGLE-----TELQSQLAKLSLE 387

281 KAFDECKSNAILKKILDNDEYKINFRWVNEVTCANTKFEALNDLIISDCEKKGIKI 340
388 ASLAETEGRYCVQLSQISQISALEOLOQIRAEETECQNAEYQQLLD-----IKT 437
341 N-RDVISSYKLLSTIYIVGAGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSTLTGN 399
438 RLENEIOTYRSLLE-----GEGSSSGGGGRRG-----GS 467
400 GGTESGGTAGTTTSSGTAGTGTGTTTSSGAASGKAGTGTAGTTTSSGAGSKAGTGTGTS 459
468 GGSYGGSSG-----GGSYGGSSGGGSGYGGSSG--GGGSYGGSSGGSGGGSGGGYG 520
460 GTTSSGTGAGGAGSGGSPSHASNA 484
521 GGSSSGAGGGGGSGGGYGGSSS 545

RESULT 10
VLPB_MYCHR STANDARD; PRT; 174 AA.
AC P29229;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Variant surface antigen B precursor (VLPB prolipoprotein).
GN VLPB.
OS Mycoplasma hyorhinis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2100;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SK76;
RX MEDLINE=92097525; PubMed=1721868;
RA Yogen D., Rosengarten R., Watson-McKown R., Wise K.S.;
RT "Molecular basis of Mycoplasma surface antigenic variation: a novel
   set of divergent genes undergo spontaneous mutation of periodic
   coding regions and 5' regulatory sequences.";
RL EMBL J. 10:4069-4079(1991).
CC -!- FUNCTION: RESPONSIBLE FOR THE ANTIGENIC DIVERSITY FOR HOST
   ADAPTATION.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
   (Probable).
CC -!- MISCELLANEOUS: THE NUMBERS OF REPEATS CAN VARY AND IS ONE OF THE
   BASIS OF THE ANTIGENIC DIVERSITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (see http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62936; CAA44709.1; ALT_SEQ.
DR PIR; S18654; S18654.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Antigen; Membrane; Lipoprotein; Repeat; Signal.
FT SIGNAL 1 29 PROBABLE.
FT CHAIN 30 174 VARIANT SURFACE ANTIGEN B.
FT LIPID 30 30 N-ACYL DIGLYCERIDE (PROBABLE).
FT DOMAIN 88 171 7 X 12 AA TANDEM REPEATS.
FT REPEAT 88 99 1.
FT REPEAT 100 111 2.
FT REPEAT 112 123 3.
FT REPEAT 124 135 4.
FT REPEAT 136 147 5.
FT REPEAT 148 159 6.
FT REPEAT 160 171 7.
SQ SEQUENCE 174 AA; 16145 MW; 28F4C9ECA85585D7 CRC64;

Query Match 7.4%; Score 190.5; DB 1; Length 174;
Best Local Similarity 35.6%; Pred. No. 0.00058;
Matches 53; Conservative 25; Mismatches 58; Indels 13; Gaps 5;
```

```
Query Match      7.7%; Score 196.5; DB 1; Length 1224;
Best Local Similarity 21.7%; Pred. No. 0.0029;
Matches 136; Conservative 69; Mismatches 220; Indels 203; Gaps 24;

QY 25 GEVTSNFRKISKEYEYETEL-----AKEHCCKKCVNVDNEDNNLKIIYAKQKSVVT 78
DB 320 GLKSGGFGVIGRPVSTPEPFLGLTFREAPPEEARPDNYM-VSNGTNMLLVICATPKSSYK 378
QY 79 TPADV-AGVSDGFFIRGQILGAVGVNE-----QNTVGMSELEQIKNELYSFSNE 128
DB 379 VPDEILSQSKPFAIRHTATGIIISHVDSAAVSALGYLPQDILGRSIMDEYHHEDLSVYKE 438
QY 129 IYHTISQ---ISNSFLIMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKILRGARSEGN 185
DB 439 TYETVMKKGOTAGASCPCPYRFLQNGCYVL-LETEWTSFVNPSRKLFEVVG-----HH 493
QY 186 NMFQ-----EALIRFNASSEEMVNAASYLSAALFRYKFEDELEFKANDNFRDGD 237
DB 494 RVFGGPKQCNVFEAPTCKLKISEE-----AQSRNTRIKEDIVKRLAETVSRPS- 542
QY 238 YDPDYINTRKELY-----ILASVLDGL-----DLIMERLIENFSDVNNTDDIKKA--- 282
DB 543 -----DTVKQSVRRCOALASFMETLMDDEVSRADLKLELPHENELTVSERDSVLMGELS 596
QY 283 -----FDECKSNAILKKILDNDEYKINFRMENV-----TCA 318
DB 597 PHHDYDYSKST---ETPPSYNQLNYENLLRFFNSKPVTAPEALDPPKTEPPEPRGTCV 653
QY 319 NTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVGAGVEATVVSATNSGTE 378
DB 654 SGASGPMSP-VHEGSGGSGSSGNFTTASNIHMSSVTNTSIAGTG-----GTGRTGTGTG 705
QY 379 SG-GAGSGTGTSV-SATSLTLTNG-GTESGAGTAGTTSSGTEAGGTSGTTTSS----- 428
DB 706 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 765
QY 429 ----- 428
DB 766 PPVLTESLLNKHNDMEKFMKKHRESRGTGKSKSANDTLKMWLGYSGPGHIGRGG 825
QY 429 -----GAASGAGGTGTAGTTTSSGAGSDRAGTGTSGT- 461
DB 826 SHSWEAGANKPKOQLTGLTDAIKGAAGSAGGAVGTGGVGGAGV-AGGSGSGTGVAGTP 884
QY 462 ----TTSSGTGA-GGAGSGSGSGHASNA 484
DB 885 EGRATTSTGTGTPGAGGGGGGAGAAAA 912

RESULT 8
SWPL_ENCCU
ID SWPL_ENCCU STANDARD; PRT; 450 AA.
AC Q9X2V1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Spore wall protein 1 precursor.
GN SWPL
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RA Bohne W., Ferguson D.J.P., Kohler K., Gross U.;
RT "Molecular characterisation of a developmentally expressed spore wall
RT protein from the human microsporidian Encephalitozoon cuniculi.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SPORE WALL COMPONENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/

Query Match      7.5%; Score 193.5; DB 1; Length 450;
Best Local Similarity 28.7%; Pred. No. 0.0013;
Matches 51; Conservative 40; Mismatches 64; Indels 23; Gaps 5;

QY 315 VTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVGAGVEATVVSATS 374
DB 287 ITAITNALQANKNFVFTTQTNLQTDVQNALITALTLTTS-----TTSTERTQ 340
QY 375 -----NOTESGAGSGTGTSVSTATSLTNGTGESGAGTAGTTTSSGTEAGGTSGTT 425
DB 341 FANSEIGALTGRIFGSGSGSGS-SGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSS 399
QY 426 TSSGAASGKACTAGTTTSSGAGSDKAGNCTGTTTSSGAGGAGSGSGSGHASN 483
DB 400 GGS-SGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSS 450

RESULT 9
KICJ_MOUSE
ID KICJ_MOUSE STANDARD; PRT; 569 AA.
AC P02535; P08731;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type I cytoskeletal 10 (Cytokeratin 10) (56 kDa cytoke(ratin)
DE (Keratin, type I cytoskeletal 59 kDa).
GN KRT10 OR KRT1-10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85207552; PubMed=2581944;
RA Krieg T.M., Schafer M.P., Cheng C.K., Filpula D., Flaherty P.,
RA Steinert P.M., Roop D.R.;
RT "Organization of a type I keratin gene. Evidence for evolution of
RT intermediate filaments from a common ancestral gene.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=83192464; PubMed=6188955;
RA Steinert P.M., Rice R.H., Roop D.R., Trus B.L., Steven A.C.;
RT "Complete amino acid sequence of a mouse epidermal keratin subunit
RT and implications for the structure of intermediate filaments.";
RN [3]
RP SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
RX Nature 302:794-800(1983).
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC MICROFILBRILLAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
CC (40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
```







DT 20-MAR-1987 (Rel. 04, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Keratin, type II cytoskeletal 1 (Cytokeratin 1) (K1) (CK 1) (67 kDa  
 DE cyokeratin) (Hair alpha protein).  
 GN KRT1 OR KRTA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo;  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=85166239; PubMed=2580302;  
 RA Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;  
 RA "Structure of a gene for the human epidermal 67-kDa keratin";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RA Whitlock N.V., Eady R.A., McGrath J.A.;  
 RA "Genomic organization of the human keratin 1 gene";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RA Hatsell S.J., Eady R.A., Wennerstrand L., Dopping-Hepenstal P.,  
 RA Leigh I.M., Munro C., Kelsell D.P.;  
 RA "Novel splice site mutation in keratin 1 underlies mild epidermolytic  
 RA palmoplantar keratoderma in three kindreds";  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RN PRELIMINARY SEQUENCE OF 151-643 FROM N.A.  
 RX MEDLINE=85207740; PubMed=2581964;  
 RA Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,  
 RA Roop D.R.;  
 RA "Amino acid sequences of mouse and human epidermal type II keratins  
 RA of Mr 67,000 provide a systematic basis for the structural and  
 RA functional diversity of the end domains of keratin intermediate  
 RA filament subunits";  
 RL J. Biol. Chem. 260:7142-7149(1985).  
 RN [5]  
 RN REVISIONS, AND VARIANT EHK PRO-160.  
 RX MEDLINE=92386601; PubMed=1381288;  
 RA Chipev C.C., Korge B.P., Markova N., Bale S.J., Digiovanna J.J.,  
 RA Compton J.G., Steinert P.M.;  
 RA "A leucine-->proline mutation in the H1 subdomain of keratin 1  
 RA causes epidermolytic hyperkeratosis";  
 RL Cell 70:821-828(1992).  
 RN [6]  
 RN VARIANTS EHK GLY-154; SER-187 AND PRO-192.  
 RX MEDLINE=94117869; PubMed=7507151;  
 RA Yang J.-M., Chipev C.C., Digiovanna J.J., Bale S.J., Marekov L.N.,  
 RA Steinert P.M., Compton J.G.;  
 RA "Mutations in the H1 and I2 domains in the keratin 1 gene in  
 RA epidermolytic hyperkeratosis";  
 RL J. Invest. Dermatol. 102:17-23(1994).  
 RN [7]  
 RN VARIANTS EHK PRO-185 AND SER-187.  
 RX MEDLINE=94117870; PubMed=7507152;  
 RA McLean W.H.I., Eady R.A., Dopping-Hepenstal P.J., McMillan J.R.,  
 RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,  
 RA Morley S.M.;  
 RA "Mutations in the rod 1A domain of keratins 1 and 10 in bullous  
 RA congenital ichthyosiform erythroderma (BCIE)";  
 RL J. Invest. Dermatol. 102:24-30(1994).  
 RN [8]  
 RN VARIANT EHK GLN-489.  
 RX MEDLINE=92376531; PubMed=1380725;  
 RA Rothnagel J.A., Dagny A.M., Dempsey L.D., Longley M.A.,  
 RA Greenhalgh D.A., Gagne T.A., Huber E., Frenk E., Hohl D., Roop D.R.;  
 RA "Mutations in the rod domains of keratins 1 and 10 in epidermolytic  
 RA hyperkeratosis";  
 RL Science 257:1128-1130(1992).  
 RN [9]  
 RN VARIANT ALLELE 1B.  
 RX MEDLINE=93107743; PubMed=1281859;  
 RA Korge B.P., Compton J.G., Steinert P.M., Mischke D.;  
 RA "The two size alleles of human keratin 1 are due to a deletion in the  
 RA glycine-rich carboxyl-terminal V2 subdomain";  
 RL J. Invest. Dermatol. 99:697-702(1992).  
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
 CC KERATIN 1 IS GENERALLY ASSOCIATED WITH KERATIN 10.  
 CC TISSUE SPECIFICITY: THE SOURCE OF THIS PROTEIN IS NEONATAL  
 CC FORESKIN. THE 67-KDA TYPE II KERATINS ARE EXPRESSED IN TERMINALLY  
 CC DIFFERENTIATING EPIDERMIS.  
 CC POLYMORPHISM: THERE ARE TWO SIZE VARIANTS OF KRT1, TERMED 1A AND  
 CC 1B WITH ALLELIC FREQUENCIES OF 0.61 AND 0.39. 1B LACKS 7 LACKS  
 CC 7 RESIDUES COMPARED TO 1A.  
 CC DISEASE: DEFECTS IN KRT1 ARE A CAUSE OF EPIDERMOLYTIC  
 CC HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL  
 CC ICHTHYOSIFORM ERYTHRODERMA (BIE)); A HEREDITARY SKIN DISORDER  
 CC CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM  
 CC CORNEUM.  
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND  
 CC MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II  
 CC (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF237621; AAF60327.1;  
 CC EMBL; M98776; AAB47721.1;  
 CC EMBL; AF304164; AAG41947.1;  
 CC EMBL; M10938; AAA36153.1; ALT\_SEQ.  
 CC PIR; A02950; KRH2.  
 CC PIR; A22940; A22940.  
 CC Aarhus/Ghent-2DPAGE; 4606; NEPHGE.  
 CC Genus; HGNC:6412; KRT1.  
 CC MIM; 139350;  
 CC MIM; 113800;  
 CC InterPro; IPR001664; IF.  
 CC InterPro; IPR003054; Keratin\_II.  
 CC Pfam; PF00038; filament; 3.  
 CC PRINTS; PR01276; TYPE2KERATIN.  
 CC PROSITE; PS00226; IF; 1.  
 CC Intermediate filament; Coiled coil; Keratin; Disease mutation;  
 KW Polymorphism; Phosphorylation.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT DOMAIN 1 178 HEAD.  
 FT DOMAIN 179 488 ROD.  
 FT DOMAIN 489 643 TAIL.  
 FT DOMAIN 179 214 COIL 1A.  
 FT DOMAIN 215 233 LINKER 1.  
 FT DOMAIN 234 325 COIL 1B.  
 FT DOMAIN 326 349 LINKER 12.  
 FT DOMAIN 350 488 COIL 2.  
 FT SITE 432 432 STUTTER.  
 FT DOMAIN 1 150 GLY/PHE/SER-RICH.  
 FT DOMAIN 501 640 GLY/SER-RICH.  
 FT MOD\_RES 65 65 PHOSPHORYLATION (BY SIMILARITY).  
 FT VARIANT 154 154 V -> G (IN EHK).  
 FT VARIANT 160 160 /FTID-VAR\_003853.  
 FT VARIANT 160 160 L -> P (IN EHK).  
 FT VARIANT 185 185 /FTID-VAR\_003854.  
 FT VARIANT 185 185 S -> P (IN EHK).  
 FT VARIANT 187 187 /FTID-VAR\_003855.  
 FT VARIANT 187 187 N -> S (IN EHK).  
 FT VARIANT 192 192 /FTID-VAR\_003856.  
 FT VARIANT 192 192 S -> P (IN EHK).  
 FT VARIANT 311 311 /FTID-VAR\_003857.  
 FT VARIANT 311 311 I -> V.  
 FT VARIANT 311 311 /FTID-VAR\_003858.



RN  
 RP VARIANTS IBS ASP-493 AND LYS-493.  
 RX MEDLINE=95038833; PubMed=7524919;  
 RA Pithnagel J.A., Traupe H., Wojcik S., Huber M., Hohl D.,  
 RA Pittelkow M.R., Saeki H., Ishibashi Y., Roop D.R.;  
 RT "Mutations in the rod domain of keratin 2e in patients with ichthyosis  
 RT bullosa of Siemens.;"  
 RL Nat. Genet. 7:485-490(1994).  
 RN [6]  
 RN VARIANT IBS PRO-485.  
 RP MEDLINE=97348962; PubMed=9204966;  
 RX Yang J.-M., Lee S., Bang H.-D., Kim W.-S., Lee E.-S., Steinert P.M.;  
 RA "A novel threonine-to-proline mutation at the end of 2B rod domain in  
 RA the keratin 2e chain in ichthyosis bullosa of Siemens.;"  
 RT J. Invest. Dermatol. 109:116-118(1997).  
 RN [7]  
 RN VARIANT IBS ASN-188.  
 RP MEDLINE=99181809; PubMed=10084318;  
 RX Arin M.J., Longley M.A., Epstein E.H. Jr., Scott G., Goldsmith L.A.,  
 RA Rothnagel J.A., Roop D.R.;  
 RA "A novel mutation in the 1A domain of keratin 2e in ichthyosis bullosa  
 RT of Siemens.;"  
 RL J. Invest. Dermatol. 112:380-382(1999).  
 RN [8]  
 RN VARIANT IBS ASP-192.  
 RP MEDLINE=20087079; PubMed=10620137;  
 RX Takizawa Y., Akiyama M., Nagashima M., Shimizu H.;  
 RA "A novel asparagine-->aspartic acid mutation in the rod 1A domain in  
 RA keratin 2e in a Japanese family with ichthyosis bullosa of Siemens.;"  
 RL J. Invest. Dermatol. 114:193-195(2000).  
 CC -1- FUNCTION: PROBABLY CONTRIBUTES TO TERMINAL CORNIFICATION.  
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
 CC -1- TISSUE SPECIFICITY: IN THE SUPRABASAL LAYERS OF EPIDERMAL  
 CC METAPLASTAS AND CARCINOMAS.  
 CC -1- DEVELOPMENTAL STAGE: SYNTHESIZED DURING MATURATION OF EPIDERMAL  
 CC KERATINOCYTES.  
 CC -1- DISEASE: DEFECTS IN KRT2A ARE A CAUSE OF ICHTHYOSIS BULLOSA OF  
 CC SIEMENS (IBS), A RARE AUTOSOMAL DOMINANT DISORDER DISPLAYING A  
 CC TYPE OF EPIDERMOLYTIC HYPERKERATOSIS CHARACTERIZED BY EXTENSIVE  
 CC BLISTERING FROM BIRTH. HYPERKERATOSIS AND SHEDDING OF THE OUTER  
 CC LAYERS OF THE EPIDERMIS (MOLTING) ARE OBSERVED IN LATER WEEKS.  
 CC -1- MUSCLEFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II  
 CC (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K9].  
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M99061; AAC83410.1; --  
 DR EMBL; AF019084; AAB81946.1; --  
 DR PIR; A44861; A44861.  
 DR Genew; HGNC:6439; KRT2A.  
 DR MIM; 600194; --  
 DR MIM; 146800; --  
 DR InterPro; IPR001664; IF.  
 DR InterPro; IPR003054; Keratin\_II.  
 DR Pfam; PF00038; filament; 1.  
 DR PRINTS; PR01276; TYPE2KERATIN.  
 DR PROSITE; PS00226; IF: 1.  
 KW Intermediate filament; Coiled coil; Keratin; Disease mutation;  
 KW Phosphorylation.  
 FT DOMAIN 1 183 HEAD.  
 FT DOMAIN 184 493 ROD.  
 FT DOMAIN 494 645 TAIL.  
 FT DOMAIN 184 219 COIL 1A.  
 FT DOMAIN 220 238 LINKER 1.

FT	DOMAIN	239	330	COIL 1B.
FT	DOMAIN	331	354	LINKER 12.
FT	DOMAIN	355	493	COIL 2.
FT	MOD_RES	62	62	PHOSPHORYLATION (BY SIMILARITY).
FT	VARIANT	187	187	Q -> P (IN IBS).
FT	VARIANT	188	188	/FTid=VAR_003865.
FT	VARIANT	192	192	I -> N (IN IBS).
FT	VARIANT	192	192	/FTid=VAR_010514.
FT	VARIANT	192	192	N -> D (IN IBS).
FT	VARIANT	192	192	N -> Y (IN IBS).
FT	VARIANT	482	482	/FTid=VAR_009185.
FT	VARIANT	485	485	E -> K (IN IBS).
FT	VARIANT	485	485	/FTid=VAR_009186.
FT	VARIANT	490	490	T -> P (IN IBS).
FT	VARIANT	490	490	L -> P (IN IBS).
FT	VARIANT	493	493	/FTid=VAR_010516.
FT	VARIANT	493	493	E -> D (IN IBS).
FT	VARIANT	493	493	/FTid=VAR_003866.
FT	VARIANT	493	493	E -> K (IN IBS).
FT	VARIANT	493	493	/FTid=VAR_003867.
SEQ	SEQUENCE	645 AA;	65865 MW;	EE025A173E33409A CRC64;

Query Match 8.58; Score 217; DB 1; Length 645;  
 Best Local Similarity 22.1%; Pred. No. 0.00011;  
 Matches 120; Conservative 100; Mismatches 200; Indels 122; Gaps 26;

QY	11	GIHVVYIDGSLASGEVTSNFRYSKEVEYEHTELAKEHCKEKCVCN-----VDNI---	61
DB	155	GIHEVSVNQSL-----QPLNVKVDPEIQNVKQAEQIKTLNKKASFIDKVRFL	205
QY	62	EDNNLKIYAK-----QFKSVVTPPADVAGVSDGF-----FIRGNLGAQVSVNEQPTVVG	111
DB	206	EQQNOVLOTWELLQOMNVGTRPINLEPIFOGYIDSLKRYLDGLTAERTSQNSLNN---	262
QY	112	MSLEQFTKNELYSFNSNIYHTISQISNSFLIMMSDAIVKHDN-YILKKKEGCEQIYNY	170
DB	263	--MODLYEDYKKYDEINERTAAE--NDFVTLKKDV-----DNAYMIKVELQSKVOLLNQ	314
QY	171	E-EFIEKLGRAR-----SEGNNMFOEALIRFRNASSEEMVNAASYLSA	212
DB	315	EIEFLKVLVDAEISQIHQSVTDTNVILSDMSNRNLDLSIAAEVKAQYEEIAQSRKEAE	374
QY	213	ALFRYKFFDELFFKANDNFGRDDGYDFDYINTKKELVILASVLDGLIMERLIENFSD	272
DB	375	ALYHSK-----YEELQVTVGR-----HGDLSLKEIKI---EISELNRVIQRLQGEIAH	418
QY	273	VNNTDDIKKAFDECKSNAILLKKILNDDEYKINFRMVNVTCAKTKFEALND----	327
DB	419	V-----KK---QCKN-----VQDAIAADAEQGEHALDKARKL---NDLEALQQAEDL	462
QY	328	--LIISDCEKKGKINRDV-ISSYKLLI-----STITYIVGAGVEAVTVSVSA	372
DB	463	ARLLRDYQELMNVKALDVELATYRKLLLEGECRMSGDLSSNVTVSVTSSTISSNVASKA	522
QY	373	TSNGTESGGAGSGTGTSVSATSTLTNGCTEGSGTAGTTTSSGTEAGTSTTTSSGAAS	432
DB	523	AFGG--SGGRGSSGGYSSGSSYSGGRGSGRG--SGGGGYSISGGYSGGGSGGSGRY	579
QY	433	GKAGTGTAGTTTSEAGS-----DKAGTGTCTTTSSGTCAGCAGSG--GPSCHASNA	484
DB	580	GSGG-GSKGGSISGGYSGGSGGKHSKSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG	638
QY	485	KI 486	
DB	639	SV 640	

RESULT 4  
 K2CL\_HUMAN  
 ID K2CL\_HUMAN STANDARD; PRT; 643 AA.  
 AC P04264; Q14720; Q9H298;

CC MICROFIBRILLAR KERATIN: I (ACIDIC; 40-55 kDa) [K9 TO K20] AND II  
 CC (NEUTRAL TO BASIC; 56-70 kDa) [K1 TO K8].  
 CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
 CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN EXTENSIVELY IN  
 CC POSITIONS 513 TO 555.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X14487; CAA32649.1; -;  
 CC EMBL; M19156; AAS59468.1; -;  
 CC EMBL; M77663; AAS59199.1; -;  
 CC EMBL; L20218; AAS59438.1; -;  
 CC EMBL; L20219; AAS59439.1; -;  
 CC PIR; S02158; KRHUO.  
 CC SWISS-2DPAGE; P13645; HUMAN.  
 CC Aarbus/Ghent-2DPAGE; 7405; IEF.  
 CC Genew; HGNC:6413; KRT10.  
 CC MIM; 148080; -;  
 CC MIM; 113800; -;  
 CC InterPro; IPR001664; IF.  
 CC InterPro; IPR002957; Keratin\_I.  
 CC Pfam; PF00038; filament; 1.  
 CC PRINTS; PRO1248; TYPE1KERATIN.  
 CC PROSITE; PS00226; IF; 1.  
 CC Intermediate filament; Coiled coil; Keratin; Disease mutation;  
 CC Polymorphism.  
 CC FT DOMAIN 1 145 HEAD.  
 CC FT DOMAIN 146 456 ROD.  
 CC FT DOMAIN 457 593 TAIL.  
 CC FT DOMAIN 146 181 COIL 1A.  
 CC FT DOMAIN 182 202 LINKER 1.  
 CC FT DOMAIN 203 294 COIL 1B.  
 CC FT DOMAIN 295 317 LINKER 12.  
 CC FT DOMAIN 318 456 COIL 2.  
 CC FT DOMAIN 451 590 GLY/PHE/SER-RICH.  
 CC FT DOMAIN 126 126 GLY/SER-RICH.  
 CC FT VARIANT 150 150 FTID-VAR\_010505.  
 CC FT VARIANT 150 150 M -> R (IN EHK).  
 CC FT VARIANT 150 150 FTID-VAR\_010506.  
 CC FT VARIANT 154 154 M -> T (IN EHK).  
 CC FT VARIANT 154 154 N -> H (IN EHK).  
 CC FT VARIANT 156 156 FTID-VAR\_003826.  
 CC FT VARIANT 156 156 R -> H (IN EHK).  
 CC FT VARIANT 156 156 FTID-VAR\_003827.  
 CC FT VARIANT 156 156 R -> C (IN EHK).  
 CC FT VARIANT 156 156 FTID-VAR\_003828.  
 CC FT VARIANT 156 156 R -> P (IN EHK).  
 CC FT VARIANT 156 156 FTID-VAR\_003829.  
 CC FT VARIANT 156 156 R -> S (IN EHK).  
 CC FT FTID-VAR\_003830.

Query Match 8.5%; Score 217.5; DB 1; Length 593;  
 Best Local Similarity 21.8%; Pred. NO. 9.5e-05;  
 Matches 111; Conservative 82; Mismatches 207; Indels 109; Gaps 19;  
 QY 18 DGSLLASCEVTS---NFR---YISKEYEYETELAKEHCCKVCVNDIEDNNLKIYA 70  
 DB 138 DGLLSGNETKMTQNLNDRSLASLDKVRALAESNYLEGTIKWEYKHGNSHOGEPRDYS 197  
 QY 71 KQFKS-----VVTTPADVAGV-----SDGFIRQNLGAVGVNQPNTVGM 113  
 DB 198 KYTKTIDDLKQNLNLTIDNANILLOIDNARLAADDFLKVEN-----EVA 243  
 QY 114 LEQFTKNELYSFSNIYHTISS-QISNSFLIMSDAIVKHNDNYILKKEGECQIYN 172

DB 244 LRQSVADI-----NGLRRVLDELTLTKADLEMQIESLELAYLKK-----NHEE 289  
 QY 173 FIEKLGRARSGNNMFQEA-----LIRFRNASSEEMVNAASLYSAALFRYKEFDEL 224  
 DB 290 EMKDLRVNVTGVDVNVNNAAPGVDLTQLNNMRSOYEQLABQNRKDAENFEK--SKEL 347  
 QY 225 FKANDNFRDGDYDFDYINTKKELVILASVLDGLDLMERKLIERFSDVNNVTDIKKAFD 284  
 DB 348 TTEIDNNIEQISSY-----KSEITELRRNVQALE-----IELQSLALKQSLEASLA 394  
 QY 285 ECKSNAILKILNDDEDYKINFRMNVNVCANTKPEALNDLIISDCEKKGKIKIN-RD 343  
 DB 395 ETEGRYCVQLSQIOAQISALEEQLOQIRAEQTECONTEYQQLLD-----IKIRLEN 444  
 QY 344 VISSYKLLITITIVGAGVEAVTVSVTSATNSCTESGAGSGTCTSVSATS---TLTGNG 400  
 DB 445 EIQTYSLLE-----GEG-----SSGGGGRGGGSGGGYGGGSGGGYGGGGH 492  
 QY 401 GTESGGTAGTTTSSGTGAGGTSGTTTSSGAASKAGTGAGTTTSSSEGAGSDKA-----G 455  
 DB 493 GSSGGYGGGSGGGSGGGYGGGSGGGHGGSSGGHGGSSGGYGGGSGGGGGGGY 552  
 QY 456 TGTSGTTTSSGTGAGGAGSGGSGGSHASNA 484  
 DB 553 GSSGGGSGGGYGGGSGGSGG--CHKSSS 579

RESULT 3  
 K22E\_HUMAN  
 ID K22E\_HUMAN STANDARD; PRT; 645 AA.  
 AC P35908;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Keratin, type II cytoskeletal 2 epidermal (Cytokeratin 2e) (K2e)  
 DE (CK 2e).  
 GN KRT2A OR KRT2E.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 ON NCBI\_TaxID=9606;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RP TISSUE=Thigh epidermis;  
 RX MEDLINE=92380238; PubMed=1380918;  
 RA Collin C., Moll R., Kubicka S., Ouhayoun J.-P., Franke W.W.;  
 RT "Characterization of human cytokeatin 2, an epidermal cytoskeletal  
 RL protein synthesized late during differentiation.";  
 RL Exp. Cell Res. 202:132-141(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A. AND VARIANTS IBS TYR-192 AND LYS-482.  
 RX MEDLINE=99019489; PubMed=9804344;  
 RA Smith F.J.D., Maingi C., Covello S.P., Higgins C., Schmidt M.,  
 RA Lane E.B., Uitto J., Leigh I.M., McLean W.H.I.;  
 RT "Genomic organization and fine mapping of the keratin 2e gene (KRT2E):  
 RT K2e VI domain polymorphism and novel mutations in ichthyosis bullosa  
 RT of Siemens.";  
 RL J. Invest. Dermatol. 111:817-821(1998).  
 RN [3]  
 RP VARIANT IBS LYS-493.  
 RX MEDLINE=94358459; PubMed=7521371;  
 RA McLean W.H.I., Morley S.M., Lane E.B., Eady R.A.J., Griffiths W.A.D.,  
 RA Paige D.G., Harper J.I., Higgins C., Leigh I.M.;  
 RT "Ichthyosis bullosa of Siemens--a disease involving keratin 2e.";  
 RL J. Invest. Dermatol. 103:277-281(1994).  
 RN [4]  
 RP VARIANTS IBS PRO-187; PRO-490 AND LYS-493.  
 RX MEDLINE=94358461; PubMed=8077693;  
 RA Kremer H., Zeeuwen P., McLean W.H.I., Mariman E.C.M., Lane E.B.,  
 RA van de Kerkhof P.C.M., Ropers H.-H., Steijlen P.M.;  
 RT "Ichthyosis bullosa of Siemens is caused by mutations in the keratin  
 RT 2e gene.";  
 RL J. Invest. Dermatol. 103:286-289(1994).

RESULT 2  
 K1CJ\_HUMAN ID K1CJ\_HUMAN STANDARD; PRT; 593 AA.  
 AC P13645; 1990 (Rel. 13, Created)  
 DT 01-JAN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Keratin, type I cytoskeletal 10 (Cytokeratin 10) (K10) (CK 10).  
 GN KRT10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89125611; PubMed=2464696;  
 RT Rieger M., Franke W.W.;  
 RT "Identification of an orthologous mammalian cytokeratin gene. High  
 RT degree of intron sequence conservation during evolution of human  
 RT cytokeratin 10.";  
 RL J. Mol. Biol. 204:841-856(1988).  
 RN [2]  
 RP SEQUENCE OF 130-593 FROM N.A.  
 RX MEDLINE=88122104; PubMed=2448602;  
 RA Darmon M.Y., Senat A., Darmon M.C., Vasseur M.;  
 RA "Sequence of a cDNA encoding human keratin No 10 selected according  
 RT to structural homologies of keratins and their tissue-specific  
 RT expression.";  
 RL Mol. Biol. Rep. 12:277-283(1987).  
 RN [3]  
 RP SEQUENCE OF 197-593 FROM N.A.  
 RX MEDLINE=92339897; PubMed=1378806;  
 RA Tkachenko A.V., Buchman V.L., Bliskovsky V.V., Shvets Y.P.,  
 RA Kiselev L.L.;  
 RT "Exons I and VII of the gene (Ker10) encoding human keratin 10  
 RT undergo structural rearrangements within repeats.";  
 RL Gene 116:245-251(1992).  
 RN [4]  
 RP SEQUENCE OF 180-184 AND 577-589..  
 RC TISSUE=Keratinocytes;  
 RX MEDLINE=93162043; PubMed=1286667;  
 RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,  
 RA Vandeckerkhove J.;  
 RT "Microsequences of 145 proteins recorded in the two-dimensional gel  
 RT protein database of normal human epidermal keratinocytes.";  
 RL Electrophoresis 13:960-969(1992).  
 RN [5]  
 RP VARIANT EHK HIS-156.  
 RX MEDLINE=92386600; PubMed=1381287;  
 RA Cheng J., Syder A.J., Yu Q.-C., Letal A., Paller A.S., Fuchs E.;  
 RT "The genetic basis of epidermolytic hyperkeratosis: a disorder of  
 RT differentiation-specific epidermal keratin genes.";  
 RL Cell 70:811-819(1992).  
 RN [6]  
 RP VARIANTS.  
 RX MEDLINE=92141228; PubMed=1371013;  
 RA Korge B.P., Gan S.-Q., McBridge O.W., Mischke D., Steinert P.M.;  
 RT "Extensive size polymorphism of the human keratin 10 chain resides in  
 RT the C-terminal V2 subdomain due to variable numbers and sizes of  
 RT glycine loops.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:910-914(1992).  
 RN [7]  
 RP VARIANTS EHK HIS-156 AND SER-161.  
 RX MEDLINE=92376531; PubMed=1380725;  
 RA Rothnagel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,  
 RA Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.;  
 RT "Mutations in the rod domains of keratins 1 and 10 in epidermolytic  
 RT hyperkeratosis.";  
 RL Science 257:1128-1130(1992).  
 RN [8]  
 RP VARIANTS EHK HIS-154; CYS-156; HIS-156; ASP-160 AND GLN-442.  
 RX MEDLINE=94136477; PubMed=7508181;  
 RA Chipev C.C., Yang J.-M., Digiovanna J.J., Steinert P.M., Marekov L.,  
 RA Compton J.G., Bale S.J.;  
 RT "Preferential sites in keratin 10 that are mutated in epidermolytic  
 RT hyperkeratosis.";  
 RL Am. J. Hum. Genet. 54:179-190(1994).  
 RN [9]  
 RP VARIANTS EHK ARG-150; CYS-156 AND GLU-439, AND VARIANT SER-126.  
 RX MEDLINE=94216497; PubMed=7512983;  
 RA Syder A.J., Yu Q.-C., Paller A.S., Giudice G., Pearson R., Fuchs E.;  
 RT "Genetic mutations in the K1 and K10 genes of patients with  
 RT epidermolytic hyperkeratosis. Correlation between location and  
 RT disease severity.";  
 RL J. Clin. Invest. 93:1533-1542(1994).  
 RN [10]  
 RP VARIANT EHK ASN-160.  
 RX MEDLINE=94117868; PubMed=7507150;  
 RA Rothnagel J.A., Longley M.A., Holder R.A., Kuster W., Roop D.R.;  
 RT "Prenatal diagnosis of epidermolytic hyperkeratosis by direct gene  
 RT sequencing.";  
 RL J. Invest. Dermatol. 102:13-16(1994).  
 RN [11]  
 RP VARIANTS EHK PRO-156 AND SER-156.  
 RX MEDLINE=94117870; PubMed=7507152;  
 RA McLean W.H.I., Eady R.A.J., Dopping-Hepenstal P.J.C., McMillan J.R.,  
 RA Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,  
 RA Morley S.M.;  
 RT "Mutations in the rod 1A domain of keratins 1 and 10 in bullous  
 RT congenital ichthyosiform erythroderma (BCIE).";  
 RL J. Invest. Dermatol. 102:24-30(1994).  
 RN [12]  
 RP VARIANT EHK THR-150.  
 RX MEDLINE=95059228; PubMed=7526210;  
 RA Paller A.S., Syder A.J., Chan Y.-M., Yu Q.-C., Hutton M.E., Tadini G.,  
 RA Fuchs E.;  
 RT "Genetic and clinical mosaicism in a type of epidermal nevus.";  
 RL New Engl. J. Med. 331:1408-1415(1994).  
 RN [13]  
 RP VARIANT AEI THR-446.  
 RX MEDLINE=99072665; PubMed=9856845;  
 RA Suga Y., Duncan K.O., Heald P.W., Roop D.R.;  
 RT "A novel helix termination mutation in keratin 10 in annular  
 RT epidermolytic ichthyosis, a variant of bullous congenital  
 RT ichthyosiform erythroderma.";  
 RL J. Invest. Dermatol. 112:506-508(1999).  
 RN [14]  
 RP VARIANT EHK SER-160.  
 RX MEDLINE=99215719; PubMed=10201536;  
 RA Arin M.J., Longley M.A., Anton-Lamprecht I., Kurze G., Huber M.,  
 RA Hohl D., Rothnagel J.A., Roop D.R.;  
 RT "A novel substitution in keratin 10 in epidermolytic hyperkeratosis.";  
 RL J. Invest. Dermatol. 112:506-508(1999).  
 CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.  
 CC KERATIN 10 IS GENERALLY ASSOCIATED WITH KERATIN 1.  
 CC -1- TISSUE SPECIFICITY: SEEN IN ALL SUPRABASAL CELL LAYERS INCLUDING  
 CC STRATUM CORNEUM.  
 CC -1- POLYMORPHISM: A NUMBER OF ALLELES ARE KNOWN THAT MAINLY DIFFER IN  
 CC THE GLY-RICH REGION (POSITIONS 490-560).  
 CC -1- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF EPIDERMOLYTIC  
 CC HYPERKERATOSIS (EHK) (ALSO KNOWN AS BULLOUS CONGENITAL  
 CC ICHTHYOSIFORM ERYTHRODERMA (BCIE)); A HEREDITARY SKIN DISORDER  
 CC CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM  
 CC CORNEUM. AT BIRTH, AFFECTED INDIVIDUALS USUALLY PRESENT WITH  
 CC REDNESS, BLISTERS AND SUPERFICIAL EROSIONS DUE TO CYTOLYSIS.  
 CC WITHIN A FEW WEEKS, THE ERYTHRODERMA AND BLISTER FORMATION  
 CC DIMINISH AND HYPERKERATOSIS DEVELOP. TRANSMISSION IS AUTOSOMAL  
 CC DOMINANT, BUT MOST CASES ARE SPORADIC.  
 CC -1- DISEASE: DEFECTS IN KRT10 ARE THE CAUSE OF ANNULAR EPIDERMOLYTIC  
 CC ICHTHYOSIS (AEI), A DISTINCT PHENOTYPIC VARIANT OF EPIDERMOLYTIC  
 CC HYPERKERATOSIS. IT RESEMBLES CLINICAL AND HISTOLOGIC FEATURES OF  
 CC BOTH EPIDERMOLYTIC HYPERKERATOSIS AND ICHTHYOSIS BULLOSA OF  
 CC SIEMENS.  
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND



Search completed: July 16, 2003, 18:09:58  
Job time : 61.3761 secs

```

Db      773 NDEMEKFMILKHHRESRRTGCRKSKSANDTLKMLEYSGPGHGIKRGSGHSEWEANKPKO 832
      421 --TSGTITSSGAGSKAG-TGTAGTITSEGG-AGSDKAGTGTSGT-----TTSSGTGA-G 470
      833 QLTGLGDAIKGAGSAGGAVGTGGVGGGAGVAGGGSGTGVAGTPEGRATTTSGTGTGP 892

QY      471 GAGSGPGSGHASNA 484
      893 GAGGGGAGACAAA 906

Db

RESULT 13
KRMSEI
keratin, 59k type I cytoskeletal - mouse
N:Alternate names: 59-kDa type I keratin
C:Species: Mus musculus (house mouse)
C:Date: 13-Nov-1984 #sequence_revision 04-Dec-1986 #text_change 10-Dec-1999
C:Accession: A02940
C:Rrieg, T.M.; Schafer, M.P.; Cheng, C.K.; Filipula, D.; Flaherty, P.; Steinert, P.M.; R
J. Biol. Chem. 260, 5867-5870, 1985
A:Title: Organization of a type I keratin gene. Evidence for evolution of intermediate f
A:Reference number: A02940; MUID:85207552; PMID:2581944
A:Accession: A02940
A:Molecule type: DNA
A:Residues: 1-569 <KRI>
A:Cross-references: GB:L00193; GB:K00391; NID:g198625; PIDN:AAA39391.1; PID:g387397
A:Note: Initiator Met not shown
A:Note: the authors translated the codon GAG for residue 41 as Gly
C:Comment: Fourier analysis has identified a 7-residue repeating pattern (heptad) between
forms a stable alpha-helical coiled coil but is interrupted by three short regions with
C:Comment: Most of the introns of the gene encoding this protein are located within the
he sequence at or near the beginning of heptad repeats. Several of these sites are conse
C:Comment: The amino and carboxyl ends are rich in glycine, serine, and aromatic residue
C:Genetics: 206/3; 234/2; 286/3; 340/3; 382/3; 455/2; 568/2
A:Introns: 206/3; 234/2; 286/3; 340/3; 382/3; 455/2; 568/2
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament
F:1-143/Domain: head <HED>
F:143/Domain: El and V1 subdomains
F:144-457/Domain: rod <ROD>
F:144-178/Region: coil 1A
F:179-192/Region: linker 1
F:193-293/Region: coil 1B
F:294-309/Region: linker 12
F:310-328/Region: coil 2A
F:329-336/Region: linker 2
F:337-457/Region: coil 2B
F:395/Region: stutler
F:458-569/Domain: tail <END>
F:458-569/Region: V2 and E2 subdomains

Query Match          7.5%; Score 192; DB 1; Length 569;
Best Local Similarity 22.0%; Pred. No. 0.0053;
Matches 111; Conservative 68; Mismatches 192; Indels 134; Gaps 19;

QY      19 GSLLA-SGVVTS---NFR---YISKEYEYEHTELAKHCCKKCVNVDNIENNLIKIVAK 71
      136 GSLLSGRVYTMQNLDRLASTMVKRALESNYEELEKIKREYKHEKHNSSOREPRDYSK 195

Db

QY      72 QFKS-----YVTPPADVAGY-----SDGFFIRGQNLGAVGSVNEQPTVGMSL 114
      196 YKTIEDLKQILTLTDNANVLLQIDNARLAADDFRLKYEN-----EVTLRQSV 245

QY      115 EGFINKELYSFNEIYHHTISSQISNLSFLIMSDAIVKHDNYILTKKEGECGEOIYNYE--- 171
      246 EADI-NGLRRVLDL-----TLKSDLEMQTESNEELAYLKKHHEEMRDLQNVSTGD 298

QY      172 -----EFIEKLRGARGSEGNMFQEALIRFNASSERVMVAASYLSAALFRYKEF 220
      299 VNVENNAAPGVDTQLLNMRNQ-----YEAEKRNKDAEWFNOKS----- 341

QY      221 DDELFFKANDNFGRDDGYDFDYINTKKELVILASVLDGLDILMERLIENFSDVNNNTDIIK 280

```

```

Db      342 -KELTTEIDSNIEQMSSH-----KSEITELRRVTQGLE-----IELQSOLALKQSLE 387
      281 KAFDECKSNAIILKKKILNDDEYDKINPREMVEYTCANTKFEALNDLLIISDCEKKGIKI 340
      388 ASLAETGRYCVQLSQISOISALEEQLOQIRAEETCONAEYQOQLLD-----IKT 437

QY      341 N-RDVISSYKLLLSITTYIVGAGVEAVTVSVSATSNGTESGGAGSGTGTSVSATSTLTGN 399
      438 RLENEIQYRSILLE-----CEGSSSGGGGRRG-----GS 467

QY      400 GGTESGGTAGTTTSSGTAGTSGTTTSSGAASGAGTGTAGTCTTTSSGAGSDKAGTGT 459
      468 GGGSYGGSSG-----GGSYGGSSGGGSGYGGSSG---GGSYGGSSGGSGHGGSGGGYG 520

QY      460 GTTTSSTGTCAGCAGSGGPGSGHASNA 484
      521 GGSSSGGAGGHGGSGGGYGGGSS 545

Db

RESULT 14
S18654
variant surface antigen vlpB precursor - Mycoplasma hyorhinis
N:Alternate names: lipoprotein
C:Species: Mycoplasma hyorhinis
C:Date: 22-Nov-1993 #sequence_revision 14-Jul-1995 #text_change 21-Jul-2000
C:Accession: S18654
R:Yogev, D.; Rosengarten, R.; Watson-McKown, R.; Wise, K.S.
EMBO J. 10, 4069-4079, 1991
A:Title: Molecular basis of Mycoplasma surface antigenic variation: a novel set of di
A:Reference number: S18651; MUID:92097525; PMID:1721868
A:Accession: S18654
A:Molecule type: DNA
A:Residues: 1-174 <YOG>
A:Cross-references: EMBL:X62936; NID:g4495092; PIDN:CAA44709.1; PID:g581323
C:Genetics:
A:Gene: vlpB
A:Genetic code: SGC3
A:Start codon: GTG
C:Superfamily: glycine-rich cell wall structural protein 1
C:Keywords: surface antigen
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-174/Product: variant surface protein vlpB #status predicted <MAT>

Query Match          7.4%; Score 190.5; DB 2; Length 174;
Best Local Similarity 35.6%; Pred. No. 0.0014;
Matches 53; Conservative 25; Mismatches 58; Indels 13; Gaps 5;

QY      345 ISSYKLLLSITTYIVGAGVEAVTVSVSATSNGTESGGAGSGTGTSVSATSTLTNGGTES 404
      5 IFSKLLVYSGSLVALASIPLIAISCGQT-NTDKSQQPGSGSSSTSGGSGTCLGS-GTTT 62

QY      405 GGTAGTTTSSGTAGTSGTTTSSGAASGAGTGTAGTCTTTSSGAGSDKAGTGTGT-TT 463
      63 GGQSGTGT-----GGRSGSGSSSTTGGTGTGSDGSDGAKGTGSDSGAKGTGSDSGAKGTGSD 116

QY      464 SSGTCAGGAGS-----GGPSGHASNAKIPG 488
      117 SQDSGAKGTGSDSGAKGTGSDSQDSG 145

Db

RESULT 15
S00273
period clock protein - Acetabularia mediterranea chloroplast (fragment)
C:Species: chloroplast Acetabularia mediterranea
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 20-Jun-2000
C:Accession: S00273
R:Li-Weber, M.; de Groot, E.J.; Schweiger, H.G.
Mol. Gen. Genet. 209, 1-7, 1987
A:Title: Sequence homology
A:Reference number: S00273
A:Accession: S00273
A:Molecule type: DNA

```









Db 523 AFGG--SGRGSSGGGYSSGSSSYSGGGRSGRGG--SGGGGSGSGGYGGGGSGGRY 579  
QY 433 CKAGTGAGTTTSGEGAGS-----DKAGTGTSGTTTSGTGAGGAGSG--GPGSGHASNA 484  
Db 580 GSGG-GSKGGISGGYGGGKHSSGGSGRSGSGGGYGGGGSSSVKSGSGGEAFGS 638  
QY 485 KI 486  
Db 639 SV 640

RESULT 4  
KRH00  
keratin 10, type I, cytoskeletal - human  
N:Alternate names: cytokeratin 10  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 10-Dec-1999  
C:Accession: S02158; C38182; B38182; PC1102; S14666; S14669  
R:Rieger, M.; Franke, W.W.  
J. Mol. Biol. 204, 841-856, 1988  
A:Title: Identification of an orthologous mammalian cyto keratin gene. High degree of int  
A:Reference number: S02158; MUID:89125611; PMID:2464696  
A:Accession: S02158  
A:Molecule type: DNA  
A:Residues: 1-593 <R1E>  
A:Cross-references: EMBL:X14487; NID:G28316; PIDN:CAA32649.1; PID:G28317  
A:Experimental source: clone lambda-KH10-5  
R:Korge, B.P.; Gan, S.Q.; McBride, O.W.; Mischke, D.; Steinert, P.M.  
Proc. Natl. Acad. Sci. U.S.A. 89, 910-914, 1992  
A:Title: Extensive size polymorphism of the human keratin 10 chain resides in the C-term  
A:Reference number: A38182; MUID:92141228; PMID:1371013  
A:Accession: C38182  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 452-593 <KOR1>  
A:Cross-references: PIDN:AA821315.1; PID:G244509  
A:Note: sequence extracted from NCBI backbone (NCBIP:79427)  
A:Accession: B38182  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 452-463, 'P', 465-507, 'Y', 523-593 <KOR2>  
A:Cross-references: PIDN:AA821314.1; PID:G244508  
A:Note: sequence extracted from NCBI backbone (NCBIP:79431)  
R:Tkachenko, A.V.; Buchman, V.L.; Bliskovsky, V.V.; Shvets, Y.P.; Kisselev, L.L.  
Gene 116, 245-251, 1992  
A:Title: Exons I and VII of the gene (Ker10) encoding human keratin 10 undergo structural  
A:Reference number: PC1102; MUID:92333897; PMID:1378806  
A:Accession: PC1102  
A:Molecule type: mRNA  
A:Residues: 'G', 198-407, 'Q', 409-450, 'G', 452-486, 491-524, 534-593 <TRA>  
A:Cross-references: GB:M77663; NID:g186628; PIDN:AAA59199.1; PID:g186629  
A:Experimental source: embryonic skin, clone HK51  
R:Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.  
Mol. Biol. Rep. 12, 277-283, 1987  
A:Title: Sequence of a cDNA encoding human keratin No 10 selected according to structural  
A:Reference number: S14666; MUID:88122104; PMID:2448602  
A:Accession: S14666  
A:Molecule type: mRNA  
A:Residues: 130-278, 'YV', 281-311, 'I', 313-339, 'V', 341-373, 'R', 375-407, 'Q', 409-459, 'RS', 46  
56-579, 'P', 581-593 <DAR1>  
A:Cross-references: EMBL:M19156; NID:g186769  
R:Darmon, M.Y.; Semat, A.; Darmon, M.C.; Vasseur, M.  
submitted to the EMBL Data Library, May 1988  
A:Reference number: S14667  
A:Accession: S14669  
A:Molecule type: mRNA  
A:Residues: 130-278, 'YV', 281-311, 'I', 313-339, 'V', 341-373, 'R', 375-407, 'Q', 409-459, 'RS', 46  
56-593 <DAR2>  
A:Cross-references: EMBL:M19156; NID:g186769; PIDN:AAA59468.1; PID:g307086  
A:Note: the translated sequence in GenBank entry HUMKRT10A, release 111.0, differs from  
C:Genetics:  
A:Gene: GDB:KRT10; KPP

A:Cross-references: GDB:118828; OMIM:148080  
A:Map position: 17q12-17q21  
A:Introns: 209/3; 237/2; 289/3; 343/3; 385/3; 458/2; 592/3  
A:Note: this gene encodes variants with considerable length polymorphism  
A:Note: mutations in this gene can cause epidermolytic hyperkeratosis and keratosis p  
C:Complex: heterotetramer of two type I and two type II proteins, usually keratin 1  
C:Superfamily: cytoskeletal keratin  
C:Keywords: coiled coil; heterotetramer; intermediate filament; polymorphism  
F:1-145/Domain: head <HEA>  
F:146-456/Domain: helical rod  
F:457-593/Domain: tail <TAI>

Query Match 8.4%; Score 216.5; DB 1; Length 593;  
Best Local Similarity 21.8%; Pred. No. 0.00028;  
Matches 111; Conservative 82; Mismatches 207; Indels 109; Gaps 19;  
QY 18 DGSLLASGEVTS-----NFR---YISKEYEYEHTELAKERCCKKCVNDNIENNLUKIYA 70  
Db 138 DGGLLSGNEKVTQNLNDRLASLDKVRALAESNYLEGGIKWEYKKGSHQGEPRDYS 197  
QY 71 KQFKS-----VVTTPADVAGY-----SDGFFIRGQNLGAVGSVNEQPTVGM 113  
Db 198 KYKTIDDLKNQILNLTNNANILLQIDNARLAADDFRLKYEN-----EVA 243  
QY 114 LEQFIKNELYSFSEIYHTISS-QISNSFLIMSDAIVKHDNYILKKEGECBQIYNYEE 172  
Db 244 LRQSVEADI---NGLRRVLDELTLTKADLEMQIESITELAYLKK-----NHEE 289  
QY 173 FIEKLRGARSEGNMFOEA-----LIRFRNASSEEMVNAASYLSAALFRYKEFDEL 224  
Db 290 EMKDLRNVSTGDVNVEMNAPGVDLTQLLNMRSYQEALAEQNRKDAEAFNEK--SKEL 347  
QY 225 FKKANDNFRDGCYDFDYINTKKEVLASVLDGLDLMERLENSDVNNTDDIKKAFD 284  
Db 348 TTEIDNNIEQISSY-----KSEITELRNVALE-----LELSQALAKUSLASLA 394  
QY 285 ECKSNAILIKKKILDNEDYKINFREVMVNEVTCAATKFEALNDLIISDCEKKGIKIN-RD 343  
Db 395 ETEGRYCVQLSQIHAQISALEEQLOQIRATEQNTFYQQLLD-----IKIRLEN 444  
QY 344 VISSYKLLSTIYIVGAGVEATVSVSATSNTSGGAGSGTGTSVSATS-----TITNG 400  
Db 445 EIQTYSLLLE-----GEG-----SSGGGRGGSGGGYGGGGSGGGYGGGGH 492  
QY 401 GTESGCTAGTTTSSGTGAGTCTTTSSGAASKAGTGTAGTTTSSGAGSDKA-----G 455  
Db 493 GSSGGYGGSGGGSGGGYGGGGSGGGHGGSGGGSGGGYGGGGSGGGGGGGY 552  
QY 456 TGTSGTTTSSGTGAGGAGSGGPGSGHASNA 484  
Db 553 GSSGGSGSGGGYGGGGSSSG--GKSSS 579

RESULT 5  
T08080

probable myrosinase-binding protein - rape  
N:Alternate names: jasmonate inducible protein  
C:Species: Brassica napus (rape)  
C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 11-May-2000  
R:Geshi, N.; Brandt, A.  
C:Accession: T08080  
A:Title: Two jasmonate-inducible myrosinase-binding proteins from Brassica napus L. s  
A:Reference number: Z16340; MUID:98192006; PMID:9530873  
A:Accession: T08080  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-680 <GES>  
A:Cross-references: EMBL:Y11483; NID:e1023101; PIDN:CAA72271.1; PID:e304150  
A:Experimental source: cv. Global; isolate a4; young seedlings  
A:Note: Jasmonate inducible

Query Match 8.4%; Score 216; DB 2; Length 680;

A:Note: translation of initiator Met is not shown  
 J. Steinert, P.M.; Parry, D.A.D.; Idler, W.W.; Johnson, L.D.; Steven, A.C.; Roop, D.R.  
 J. Biol. Chem. 260, 7142-7149, 1985  
 A:Title: Amino acid sequences of mouse and human epidermal type II keratins of M-r 67,000  
 late filament subunits.  
 A:Reference number: A92535; MUID:85207740; PMID:2581964  
 A:Accession: A02950  
 A:Molecule type: mRNA  
 A:Residues: 151-183, 'K', 185-199, 'M', 201-204, 'K', 206-236, 'S', 238-239, 'R', 241-356, 'Y', 358-  
 'S', 638-643 <STE>  
 A:Cross-references: GB:M10938; NID:g186787; PIDN:AAA36153.1; PID:g386854  
 A:Experimental source: tissue neonatal foreskin  
 A:Note: The authors translated the codon CUG for residue 476 as Met  
 R:Chipev, C.C.; Korge, B.P.; Markova, N.; Bale, S.J.; DiGiovanna, J.J.; Compton, J.G.; S  
 Cell 70, 821-828, 1992  
 A:Title: A leucine--proline mutation in the H1 subdomain of keratin 1 causes epidermol  
 A:Reference number: A43342; MUID:92386601; PMID:1381288  
 A:Accession: A43342  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 144-146, 'P', 148-159, 'P', 161-183, 'K', 185-186 <CHI>  
 A:Cross-references: GB:M98776; GB:M11215; GB:M11845; GB:M11846; NID:g1843461  
 A:Note: sequence extracted from NCBI backbone (NCBIP:112784)  
 C:Comment: The cytoskeletal and microfibrillar keratins are classified into two types, b  
 atin IF protein subunit appears to be a heterotetramer of two type I and two type II pro  
 C:Comment: Keratin 1 is expressed in terminally differentiating epidermis.  
 C:Genetics:  
 A:Gene: GDB:KRT1  
 A:Cross-references: GDB:128198; OMIM:139350  
 A:Map position: 12q11-12q13  
 A:Note: defects in this gene may result in epidermolytic hyperkeratosis  
 C:Complex: heterotetramer of two type I, usually keratin 10 (see PIR:KRH00), and two type  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil; heterotetramer; intermediate filament  
 F:4-179/Domain: head <HED>  
 F:4-143/Region: E1 and V1 subdomains  
 F:14-179/Region: H1 subdomain  
 F:180-492/Domain: rod <ROD>  
 F:180-214/Region: coil 1A  
 F:215-226/Region: linker 1  
 F:227-327/Region: coil 1B  
 F:328-344/Region: linker 12  
 F:345-363/Region: coil 2A  
 F:364-371/Region: linker 2  
 F:372-492/Region: coil 2B  
 F:430/Region: stutter  
 F:493-643/Domain: tail  
 F:493-512/Region: H2 subdomain  
 F:513-643/Region: V2 and E2 subdomains

Query Match 8.6%; Score 219.5; DB 1; Length 643;  
 Best Local Similarity 23.6%; Pred. No. 0.00021;  
 Matches 125; Conservative 76; Mismatches 152; Indels 177; Gaps 28;

Qy 85 GVSDFGFRIGQNLG-----AVGSVNEQPTVGMSELEQIKNEL-----YSFSNEYI 130  
 Db 126 GFGGGGFGGGGGYGVCPSPGIGQEV--TINQSLQLPUNVEIDPEIOKVKSREREQI 183  
 Qy 131 HTISQISNSPLIMSDAIVKHDNYILKKEGEGCEQI-----YN-----YEEFIEKLRGA 180  
 Db 184 QSLNNQFA-SPIKDV--RFLQEQNQVLOTWELQVQVDTSTRTNLEPFPESFINNLRG 240  
 Qy 181 -----RSGNNMFQALIRFRNASSEEMVNAASYLSAALFRYKFFDDELFKKAN 229  
 Db 241 VDQLKSQSRDLSLKNK-QDMVEDYRN-----KYEDEINKRTN 278  
 Qy 230 DNFGRDGYDFDYINTKKEL-----VILASVLGLD-----LIMERLIENFSDVN 274  
 Db 279 -----AENEFTYIKVDGAYMTKVDLOAKLDNLQOEIDFTALYQAELSQMOTQIS 330  
 Qy 275 NTD-----DIKAFD-----ECKS-NAIILKKILDNDEYKINFRFEMNEVTCA---- 318  
 Db 331 ETNVLSMDNNRQDLDLSIAEVRQAQNDIAQSKAEASLYQSKYEEL--QITAGRHD 389

Qy 319 ---NTRFE--ALNDLI-----ISDCEKKGIKINRD----- 343  
 Db 389 SVRNSKIEISELNRIQLRSEINDNVKQISNLQOISDAEQRGENALKDNKLNLDLE 448  
 Qy 344 -----VISSYKLLIST-----ITY-IVGAGVE-----AVTVSVSA 372  
 Db 449 ALOQAKEDTRLRLRDYOELMNTKLALDLEIATYRTLLEGEESRMSECAPNVSVTSTSH 508  
 Qy 373 T--SNGTESGAGSGTGTVSATSTLTNGNGTSGGTAGTTSSTGTEAGTSGTSS-G 429  
 Db 509 TSIISGSGSGGGGGYGVSGSSYG--SGGSGYSGGGGGGGRGSGYSGGSGSGSSYG 566  
 Qy 430 AASGKAGTGTAGTTSSTGAGSGAGTGTGTGTTSSGTGAGGAGSGGSPSG 479  
 Db 567 SGGGGGGHSGYSGSSS--GGYRGGSGGGGGSGGSGGSGSGSGSGSG 613

RESULT 3  
 A44861  
 keratin, 67K type II epidermal - human  
 N:Alternate names: cytokeratin 2, CK 2; epidermal cytokeratin 2  
 C:Species: Homo sapiens (man)  
 C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 28-Apr-1995  
 C:Accession: A44861  
 R:Collin, C.; Moll, R.; Kubicka, S.; Ouhayoun, J.P.; Franke, W.W.  
 Exp. Cell Res. 202, 132-141, 1992  
 A:Title: Characterization of human cytokeratin 2, an epidermal cytoskeletal protein s  
 A:Reference number: A44861; MUID:92380238; PMID:1380918  
 A:Accession: A44861  
 A:Molecule type: mRNA  
 A:Residues: 1-645 <COL>  
 A:Cross-references: GB:S43646  
 A:Experimental source: epidermis  
 A:Note: the authors translated the codon GGC for residue 146 as Cys  
 A:Note: sequence extracted from NCBI backbone (NCBIN:112351)  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil; intermediate filament

Query Match 8.5%; Score 217; DB 2; Length 645;  
 Best Local Similarity 22.1%; Pred. No. 0.00029;  
 Matches 120; Conservative 100; Mismatches 200; Indels 122; Gaps 26;

Qy 11 GIHHYIDGSLASGEVTSNFRYISKEYEYEHTELAKHECKKCKVN-----VDNI--- 61  
 Db 155 GIHEVSVNQSL-----OPLNVKVDPEIQNKAQEREQIKTLNKKFASIDKVRFL 205  
 Qy 62 EDNNKIYAK---QFKSVVTTPADVAGVSDGF-----FIRGNLGAAGSVNEQPTVG 111  
 Db 206 EQQNOVLOTWELQNMVGTGRPINLEPIFGYIDSLKRYLDGLTAERTSQNSELNN--- 262  
 Qy 112 MSLEQFIKNELYSFNSNIYHTISQISNSPLIMSDAIVKHDN-YILKKEGEGCEQIYNY 170  
 Db 263 --MDQLVEDYKKYKYEDEINKRTAAE--NDFVTLKDV---DNAYMIKVELQSKVDLLNQ 314  
 Qy 171 E-EFIEKLRAR-----SEGNNMFQALIRFRNASSEEMVNAASYLSA 212  
 Db 315 EIEFLKYLVDAEISQIHQSVDTDNVILSDNSNRDLDSIIAEYKAQYEEIAQSRKEAE 374  
 Qy 213 ALFRYKFFDDELFKKANDNFGDDGYDFYINTKKELVILASVLGLDGLMERLIENFSD 272  
 Db 375 ALVHSK-----YEELOQTVGR-----HGDLSLKEIKI---EISELNRIQLRQGLIAH 418  
 Qy 273 VNNTDDIKKAFDECKSNAILKKILDNDEYKINFRFEMNEVTCANTKFEALND----- 327  
 Db 419 V-----KK---QCKN---VQDAIDAEQGEHALKDARKNL---NDLEALQAKEDL 462  
 Qy 328 --LIISDCEKKGIKINRDV-ISSYKLL-----STITYIVGAGVEAVTVSVSA 372  
 Db 463 ARLLRDYQELMNVKRLALDVEIATYRKLLEGEESRMSECAPNVSVTSTSSNASKA 522  
 Qy 373 TSNCTESGGAGSGTGTVSATSTLTNGCTESGGTAGTTSSTGTEAGTSGTSSGAAS 432

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	222	8.7	713	1	UMMS		period clock prote
2	219.5	8.6	643	1	KRHU2		keratin I, type II
3	217	8.5	645	2	A4861		keratin, 67K type I
4	216.5	8.4	593	1	KRHU0		keratin 10, type I
5	216	8.4	680	2	T0800		probable myosinas
6	200.5	7.8	1428	2	T08852		lustrin A - Califo
7	199	7.8	581	1	KRMS2		keratin, type II c
8	198	7.7	1079	2	B70807		hypothetical glyci
9	197.5	7.7	561	2	A31994		keratin 10, type I
10	196.5	7.7	1176	2	C36427		period clock prote
11	196.5	7.7	1218	2	A36427		period clock prote
12	196.5	7.7	1218	2	A36588		period clock prote
13	192	7.5	569	1	KPMS1		keratin, 59K type
14	190.5	7.4	174	2	S86534		variant surface an
15	188.5	7.3	174	2	S00273		period clock prote
16	188.5	7.3	570	2	S07330		keratin, epidermal
17	186	7.3	1127	2	A35018		circadian rhythm p
18	186	7.3	1489	2	F70807		hypothetical glyci
19	185	7.2	1901	2	F70806		hypothetical glyci
20	184	7.2	967	2	S68852		hypothetical prote
21	183.5	7.2	778	2	F70963		hypothetical glyci
22	183.5	7.2	1381	2	F70806		hypothetical glyci
23	182.5	7.1	1536	2	A43855		high-molecular-we
24	181	7.1	1306	2	A70934		hypothetical glyci
25	180	7.0	2329	2	T28125		hypothetical prote
26	178.5	7.0	937	2	S68135		hypally regulated
27	178	6.9	592	2	E82759		endo-1,4-beta-gluc
28	177	6.9	731	2	C70974		hypothetical glyci
29	176	6.9	463	2	B70893		hypothetical glyci

A;Residues: 1-643 <JOH>

F;41-77/Domain: EGF homology <EGF>

[illegible]





```
; ORGANISM: Babesia
US-09-737-178-85

Query Match      44.6%; Score 1143; DB 10; Length 666;
Best Local Similarity 99.1%; Pred. No. 1.8e-65;
Matches 228; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 274 NNTDDIKKAFDECKSNAILKKILDNDEYKINFRMNEVTCANTKFEALNDLIISDC 333
Db 7 HHTDDIKKAFDECKSNAILKKILDNDEYKINFRMNEVTCANTKFEALNDLIISDC 66

QY 334 EKKIKINRDVSISSYKLLSTIYIVGAGVEAVTVSVSATSNGTSGGAGSGTGTSVSAT 393
Db 67 EKKIKINRDVSISSYKLLSTIYIVGAGVEAVTVSVSATSNGTSGGAGSGTGTSVSAT 126

QY 394 STLTGNGGTSGGTAGTTTSSGTEAGTGTSSGAAAGKAGTGTAGTTTSSGAGSDK 453
Db 127 STLTGNGGTSGGTAGTTTSSGTEAGTGTSSGAAAGKAGTGTAGTTTSSGAGSDK 186

QY 454 AGTGTSTTTSSGTGAGGAGSGGPGSHASNAKIPGIMTLTLFALLTFIVN 503
Db 187 AGTGTSTTTSSGTGAGGAGSGGPGSHASNAKIPGIMTLTLFALLTFIVN 236

RESULT 12
US-09-853-079-144
; Sequence 144, Application US/09853079
; Publication No. US20030109689A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C11
; CURRENT APPLICATION NUMBER: US/09/853,079
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 144
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Babesia
US-09-853-079-144

Query Match      44.5%; Score 1141; DB 9; Length 677;
Best Local Similarity 100.0%; Pred. No. 2.5e-65;
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 TDDIKKAFDECKSNAILKKILDNDEYKINFRMNEVTCANTKFEALNDLIISDC 335
Db 450 TDDIKKAFDECKSNAILKKILDNDEYKINFRMNEVTCANTKFEALNDLIISDC 509

QY 336 KGKINRDVSISSYKLLSTIYIVGAGVEAVTVSVSATSNGTSGGAGSGTGTSVSAT 395
Db 510 KGKINRDVSISSYKLLSTIYIVGAGVEAVTVSVSATSNGTSGGAGSGTGTSVSAT 569

QY 396 LTGNGGTSGGTAGTTTSSGTEAGTGTSSGAAAGKAGTGTAGTTTSSGAGSDK 455
Db 570 LTGNGGTSGGTAGTTTSSGTEAGTGTSSGAAAGKAGTGTAGTTTSSGAGSDK 629

QY 456 TGTSGTTTSSGTGAGGAGSGGPGSHASNAKIPGIMTLTLFALLTFIVN 503
Db 630 TGTSGTTTSSGTGAGGAGSGGPGSHASNAKIPGIMTLTLFALLTFIVN 677

RESULT 14
US-09-286-488-49
; Sequence 49, Application US/09286488
; Patent No. US20020169136A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C3
; CURRENT APPLICATION NUMBER: US/09/286,488
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 49
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-286-488-49

Query Match      9.6%; Score 246.5; DB 9; Length 367;
Best Local Similarity 28.8%; Pred. No. 2.5e-08;
Matches 92; Conservative 53; Mismatches 113; Indels 61; Gaps 12;

QY 80 PADVAGVSD--GFFIRGONLGVAGSVNEQPTVGMSEIQFIKNELYSFSEIHTISSQI 137
Db 75 PSEAGGPGSEAGGSGTGTSEAGGWPSTGTGWPSEAGWSERF--GYQLLPYSRRI--VIFNEV 131
```

```
Db 472 -----PGSTDDIKKAFDECKSNAILKKKILDNDEYKINFREMVNEVTCANTKFEALN 525
QY 327 DLIISDCCKGKIKNRDVISSYKLLSTIYIVGAGVEAVTVSVSATSNGTESGAGSGT 386
Db 526 DLIISDCCKGKIKNRDVISSYKLLSTIYIVGAGVEAVTVSVSATSNGTESGAGSGT 585
QY 387 GTSVSATSLTNGGTSAGTGTSSGTEAGTGTSSGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 503
Db 586 GTSVSATSLTNGGTSAGTGTSSGTEAGTGTSSGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 702
QY 447 EGAGSDKAGTGTSSGTTSSGTGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 503
Db 646 EGAGSDKAGTGTSSGTTSSGTGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 702
RESULT 9
US-09-737-178-87
; Sequence 87, Application US/09737178
; Publication No. US20010029295A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 87
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Babesia
US-09-737-178-87
Query Match 45.2%; Score 1159.5; DB 10; Length 1132;
Best Local Similarity 58.3%; Pred. No. 3e-66;
Matches 278; Conservative 35; Mismatches 89; Indels 75; Gaps 14;
QY 59 DNI-----EDNNKIYA-KQFKSVVTPADVAGVSDGFFIRGQNLGAGVSNQPNVTGM 112
Db 269 DNITTTNEVTKDDVYALKKALCTLTHLIYHKSVDG--ISFDMLGTKQKNKSPGLGKIGT 326
QY 113 SLEQFIKNELYSFSNEYVHTISSQISNSFLIMMSD-AIVKHDNYILKKEGCEQIYNYE 171
Db 327 SMDDIIA-----PMNVLKVAYLQAEIHFLISTKYND----IFDYT 368
QY 172 -----EFIEKLRGARGSENNMFOEALIRFERNASSEEMVNAAYLSAALFRYKE 219
Db 369 IDFSKREATDGSPTDILLGNKVKESLFIEGLISDIKSHKAGVTGGTSSSLF----424
QY 220 FDDLEFKKANDNFRDGG-----YDFDYINTKK-----ELVILASVLDGLDLIMERL 266
Db 425 --DEIF-----DELNDQATIRTLVAPLDWPLISDKSLHPSLKMVW---VLPGGFIV---471
QY 267 IENFSYDNVNTDDIKKAFDECKSNAILKKKILDNDEYKINFREMVNEVTCANTKFEALN 326
Db 472 -----PGSTDDIKKAFDECKSNAILKKKILDNDEYKINFREMVNEVTCANTKFEALN 525
QY 327 DLIISDCCKGKIKNRDVISSYKLLSTIYIVGAGVEAVTVSVSATSNGTESGAGSGT 386
Db 526 DLIISDCCKGKIKNRDVISSYKLLSTIYIVGAGVEAVTVSVSATSNGTESGAGSGT 585
QY 387 GTSVSATSLTNGGTSAGTGTSSGTEAGTGTSSGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 503
Db 586 GTSVSATSLTNGGTSAGTGTSSGTEAGTGTSSGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 702
```

```
QY 447 EGAGSDKAGTGTSSGTTSSGTGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 503
Db 646 EGAGSDKAGTGTSSGTTSSGTGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 702
RESULT 10
US-09-853-079-85
; Sequence 85, Application US/09853079
; Publication No. US20030109689A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C11
; CURRENT APPLICATION NUMBER: US/09/853,079
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 85
; LENGTH: 666
; TYPE: PRT
; ORGANISM: Babesia
US-09-853-079-85
Query Match 44.6%; Score 1143; DB 9; Length 666;
Best Local Similarity 99.1%; Pred. No. 1.8e-65;
Matches 228; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 274 NNTDDIKKAFDECKSNAILKKKILDNDEYKINFREMVNEVTCANTKFEALNDLIISDC 333
Db 7 HHTDDIKKAFDECKSNAILKKKILDNDEYKINFREMVNEVTCANTKFEALNDLIISDC 66
QY 334 EKKGIKINRDVISSYKLLSTIYIVGAGVEAVTVSVSATSNGTESGAGSGTGTSVSAT 393
Db 67 EKKGIKINRDVISSYKLLSTIYIVGAGVEAVTVSVSATSNGTESGAGSGTGTSVSAT 126
QY 394 STLTGNGTESGAGTGTSSGTEAGTGTSSGAAAGKAGTGTAGTTSSGAGSDK 453
Db 127 STLTGNGTESGAGTGTSSGTEAGTGTSSGAAAGKAGTGTAGTTSSGAGSDK 186
QY 454 AGTGTGTTSSGTGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 503
Db 187 AGTGTGTTSSGTGAGGAGGSGGPGSGHASNKIPGIMTLTFLFALLFFIVN 236
RESULT 11
US-09-737-178-85
; Sequence 85, Application US/09737178
; Patent No. US20010029295A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 85
; LENGTH: 666
; TYPE: PRT
```



```
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-853-079-21

Query Match      83.0%; Score 2130; DB 9; Length 492;
Best Local Similarity 99.5%; Pred. No. 6.9e-129;
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KRNEHTDMNGIIHYIIDGSLASGEVTSNFRISKEYEYEHTELAKHCKKCKCVNVDN 60
DB 73 KRNEHTDMNGIIHYIIDGSLASGEVTSNFRISKEYEYEHTELAKHCKKCKCVNVDN 132
QY 61 IEDNNLKIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGVSNEQNTVGMSELEQFIKN 120
DB 133 IEDNNLKIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGVSNEQNTVGMSELEQFIKN 192
QY 121 ELYSFSNEIYHTTSSQISNSFLIMMSDAIVKHNDNYILKKEGEGCEQIYNYEEFIEKLRGA 180
DB 193 ELYSFSNEIYHTTSSQISNSFLIMMSDAIVKHNDNYILKKEGEGCEQIYNYEEFIEKLRGA 252
QY 181 RSEGNMFQALIRFRNASSEEMVNAASALFRYKEFDELFKANDNFGRDDGYDF 240
DB 253 RSEGNMFQALIRFRNASSEEMVNAASALFRYKEFDELFKANDNFGRDDGYDF 312
QY 241 DYINTKKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 300
DB 313 DYINTKKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 372
QY 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVVISSYKLLSTIYIVG 360
DB 373 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVVISSYKLLSTIYIVG 432
QY 361 AGVEAVTVSVSATNSGTSGGAGSGTGTSVSATSTLTNGGTSGGTAGTTTSSGTGAG 419
DB 433 AGVEAVTVSVSATNSGTSGGAGSGTGTSVSATSTLTNGGTSGGTAGTTTSSGTWFG 491

RESULT 7
US-09-737-178-21
; Sequence 21, Application US/09737178
; Patent No. US20010029295A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-737-178-21

Query Match      83.0%; Score 2130; DB 10; Length 492;
Best Local Similarity 99.5%; Pred. No. 6.9e-129;
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KRNEHTDMNGIIHYIIDGSLASGEVTSNFRISKEYEYEHTELAKHCKKCKCVNVDN 60
DB 73 KRNEHTDMNGIIHYIIDGSLASGEVTSNFRISKEYEYEHTELAKHCKKCKCVNVDN 132
QY 61 IEDNNLKIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGVSNEQNTVGMSELEQFIKN 120
DB 133 IEDNNLKIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGVSNEQNTVGMSELEQFIKN 192
QY 121 ELYSFSNEIYHTTSSQISNSFLIMMSDAIVKHNDNYILKKEGEGCEQIYNYEEFIEKLRGA 180
DB 193 ELYSFSNEIYHTTSSQISNSFLIMMSDAIVKHNDNYILKKEGEGCEQIYNYEEFIEKLRGA 252
QY 181 RSEGNMFQALIRFRNASSEEMVNAASALFRYKEFDELFKANDNFGRDDGYDF 240
DB 253 RSEGNMFQALIRFRNASSEEMVNAASALFRYKEFDELFKANDNFGRDDGYDF 312
QY 241 DYINTKKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 300
DB 313 DYINTKKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 372
QY 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVVISSYKLLSTIYIVG 360
DB 373 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVVISSYKLLSTIYIVG 432
QY 361 AGVEAVTVSVSATNSGTSGGAGSGTGTSVSATSTLTNGGTSGGTAGTTTSSGTGAG 419
DB 433 AGVEAVTVSVSATNSGTSGGAGSGTGTSVSATSTLTNGGTSGGTAGTTTSSGTWFG 491

RESULT 8
US-09-853-079-87
; Sequence 87, Application US/09853079
; Publication No. US20030109689A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C11
; CURRENT APPLICATION NUMBER: US/09/853,079
; CURRENT FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 87
; LENGTH: 1132
; TYPE: PRT
; ORGANISM: Babesia
US-09-853-079-87

Query Match      45.2%; Score 1159.5; DB 9; Length 1132;
Best Local Similarity 58.3%; Pred. No. 3e-66;
Matches 278; Conservative 35; Mismatches 89; Indels 75; Gaps 14;

QY 59 DNI-----EDNNLKIYA-KQKFSVVTTPADVAGVSDGFFIRGQNLGAVGVSNEQNTVGM 112
DB 269 DNITTTTNEVTKDDVYALKKALCTLTHLIYHSHKVDG--ISFDMLGQTKNKSPLGKIGT 326
QY 113 SLEQFIKNELYSFSNEIYHTTSSQISNSFLIMMSDAIVKHNDNYILKKEGEGCEQIYNYE 171
DB 327 SMDIIA-----MFSN-----PNMYLVKAYLQAIHIFLSTIKYND-----IFDIT 368
QY 172 -----EFIEKLRGARGSEGNMFQALIRFRNASSEEMVNAASALFRYKE 219
DB 369 IDRSKREATSGSFTDILLGNKVKESLFIIEGLISDLKSHKAGVTGGISSSLF----- 424
QY 220 FDDELFFKANDNFGRDDG-----YDFDIYINTKK-----ELVILASVLDGLDIMERL 266
DB 425 --DEIF---DELNLDQATIRTLVAPLDWPLDISKSLHPSLKMV---VLPGFIV--- 471
QY 267 IENFSDVNTDDIKKAFDECKSNAILKKKILDNDEYKINFRMNEVTCANTKFEALN 326
```

Db 421 TSGTTSSGAAGKAGTGTAGTTTSSGAGSDKAGTGTSGTTSSGTGAGGAGSGPSSH 480  
QY 481 ASNAKIPGIMTLTLLFALLTFIVN 503  
Db 481 ASNAKIPGIMTLTLLFALLTFIVN 503

## RESULT 4

US-09-853-079-209  
; Sequence 209, Application US/09853079  
; Publication No. US20030109689A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secretist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; FILE REFERENCE: 210121.426C11  
; CURRENT APPLICATION NUMBER: US/09/853,079  
; NUMBER OF SEQ ID NOS: 224  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 209  
; LENGTH: 452  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-853-079-209

Query Match 87.3%; Score 2239.5; DB 9; Length 452;  
Best Local Similarity 89.0%; Pred. No. 5.9e-136;  
Matches 445; Conservative 0; Mismatches 4; Indels 51; Gaps 1;  
QY 1 KRFNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHECHCKKCVNVDN 60  
Db 4 KRFNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHECHCKKCVNVDN 63  
QY 61 IEDNNKLIYAKQKSVVTPADVAGVSDGFFIRGONLGVAGSVNEQPTVGMSSLEQFTKN 120  
Db 64 IEDNNKLIYAKQKSVVTPADVAGVSDGFFIRGONLGVAGSVNEQPTVGMSSLEQFTKN 123  
QY 121 ELYSFSNEIYHTISSQISNSFLIMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLKRG 180  
Db 124 ELYSFSNEIYHTISSQISNSFLIMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLKRG 183  
QY 181 RSEGNMFEALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKANDNFRDGDGYDF 240  
Db 184 RSEGNMFEALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKANDNFRDGDGYDF 243  
QY 241 DYINTKKEVLIVASVLDGLDIMERLIENFSDVNNTDDIKKAFDECKSNAILLKKKILDN 300  
Db 244 DYINTKKEVLIVASVLDGLDIMERLIENFSDVNNTDDIKKAFDECKSNAILLKKKILDN 303  
QY 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLLSITTYIVG 360  
Db 304 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLLSITTYIVG 363  
QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSLTGTNGGTSGGTAGTTTSSGTEAGG 420  
Db 364 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSLTGTNGGTSGGTAGTTTSSGTEAGG 420  
QY 421 TSGTTSSGAAGKAGTGTAGTTTSSGAGSDKAGTGTSGTTSSGTGAGGAGSGPSSH 480  
Db 393 -----GTTTSSGAGSGKAGTGTSGTTTSSGTGAGGAGSGPSSH 432  
QY 481 ASNAKIPGIMTLTLLFALLTF 500  
Db 433 ASNAKIPGIMTLTLLFALLTF 452

## RESULT 5

US-09-286-488-21  
; Sequence 21, Application US/09286488  
; Patent No. US20020169136A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; FILE REFERENCE: 210121.426C3  
; CURRENT APPLICATION NUMBER: US/09/286,488  
; CURRENT FILING DATE: 1999-04-05  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 21  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-286-488-21

Query Match 83.0%; Score 2130; DB 9; Length 492;  
Best Local Similarity 99.5%; Pred. No. 6.9e-129;  
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KRFNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHECHCKKCVNVDN 60  
Db 73 KRFNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHECHCKKCVNVDN 132  
QY 61 IEDNNKLIYAKQKSVVTPADVAGVSDGFFIRGONLGVAGSVNEQPTVGMSSLEQFTKN 120  
Db 133 IEDNNKLIYAKQKSVVTPADVAGVSDGFFIRGONLGVAGSVNEQPTVGMSSLEQFTKN 192  
QY 121 ELYSFSNEIYHTISSQISNSFLIMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLKRG 180  
Db 193 ELYSFSNEIYHTISSQISNSFLIMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLKRG 252  
QY 181 RSEGNMFEALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKANDNFRDGDGYDF 240  
Db 253 RSEGNMFEALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKANDNFRDGDGYDF 312  
QY 241 DYINTKKEVLIVASVLDGLDIMERLIENFSDVNNTDDIKKAFDECKSNAILLKKKILDN 300  
Db 313 DYINTKKEVLIVASVLDGLDIMERLIENFSDVNNTDDIKKAFDECKSNAILLKKKILDN 372  
QY 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLLSITTYIVG 360  
Db 373 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLLSITTYIVG 432  
QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSLTGTNGGTSGGTAGTTTSSGTEAG 419  
Db 433 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSLTGTNGGTSGGTAGTTTSSGTEAG 491

## RESULT 6

US-09-853-079-21  
; Sequence 21, Application US/09853079  
; Publication No. US20030109689A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secretist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; FILE REFERENCE: 210121.426C11  
; CURRENT APPLICATION NUMBER: US/09/853,079

Qy	241	DYINTKELVILASVLDGLDOLIMERLIENESDVNTDDIKAFDECKSNAILILKKKIILDN	300
Db	241	DYINTKELVILASVLDGLDOLIMERLIENESDVNTDDIKAFDECKSNAILILKKKIILDN	300
Qy	301	DEDKINFREWMNEVTCA <sup>NT</sup> KFEALNDLIISDCEKKGIKINRDVISSYKLLLSITTYIVG	360
Db	301	DEDKINFREWMNEVTCA <sup>NT</sup> KFEALNDLIISDCEKKGIKINRDVISSYKLLLSITTYIVG	360
Qy	361	AGVEAVTVSVSATSN <sup>GT</sup> ESGAGSGTGTSVSATSTLTGCGTSGGTAGTTTSSGTEAGG	420
Db	361	AGVEAVTVSVSATSN <sup>GT</sup> ESGAGSGTGTSVSATSTLTGCGTSGGTAGTTTSSGTEAGG	420
Qy	421	TSGTTTSSGAASGKAGTGTA <sup>GT</sup> TSSGAGSDKAGTGTSGTTTSSGTGAGGAGSGGPPSGH	480
Db	421	TSGTTTSSGAASGKAGTGTA <sup>GT</sup> TSSGAGSDKAGTGTSGTTTSSGTGAGGAGSGGPPSGH	480
Qy	481	ASNAKIPGIMTLTLLFALLTFIVN	503
Db	481	ASNAKIPGIMTLTLLFALLTFIVN	503

RESULT 2  
US-09-853-079-52  
; Sequence 52, Application US/09853079  
; Publication No. US20030109689A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Homer, Mary  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; OF INFECTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C11  
; CURRENT APPLICATION NUMBER: US/09/853.079  
; CURRENT FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 224  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 52  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-853-079-52

Query Match	100.0%	Score 2565;	DB 9;	Length 503;
Best Local Similarity	100.0%;	Pred. No. 9.3e-157;		
Matches 503;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	KRNEHTDMGIIHHYYIDGSLASGEVTSNFRYSKEYEYEHTELAKEHCKKEKCVNVDN	60	
Db	1	KRNEHTDMGIIHHYYIDGSLASGEVTSNFRYSKEYEYEHTELAKEHCKKEKCVNVDN	60	
Qy	61	IEDNNLIYAKQKFSVYTTTADVAGVSDGFFIRQNLGAVGSVNEQPTVGMLSLEQFIKN	120	
Db	61	IEDNNLIYAKQKFSVYTTTADVAGVSDGFFIRQNLGAVGSVNEQPTVGMLSLEQFIKN	120	
Qy	121	ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNTVEEFTEKURGA	180	
Db	121	ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNTVEEFTEKURGA	180	
Qy	181	RSEGNMFQEALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKANDNFGRDDGYDF	240	
Db	181	RSEGNMFQEALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKANDNFGRDDGYDF	240	
Qy	241	DYINTKKELVTILASVLDGLDLIMERLIENFSDVNTTDDIKKAFDECKSNAILILKKIILDN	300	
Db	241	DYINTKKELVTILASVLDGLDLIMERLIENFSDVNTTDDIKKAFDECKSNAILILKKIILDN	300	
Qy	301	DEDYKINFRMVMNEVTCANTKFEALNDLIISDCBKKGIKINRDVISSYKLLLSITTVIYG	360	

```

Db      301 DEDYKINPREMNEVTCAKTFEALNDLIISDCEKKGIKINRDVVISSYKLLLSITYYIVG 360
Qy      361 AGEAVTVSVSATSNGTSESGAGSGTGTSVSATSTLTNGNGTSEGGTAGTTTSSGTEAGG 420
Db      361 AGEAVTVSVSATSNGTSESGAGSGTGTSVSATSTLTNGNGTSEGGTAGTTTSSGTEAGG 420
Qy      421 TSGTTTSSGAASGKAGTGCTAGTTTSSBAGSDKAGTGTSGTTTSSGTGAGGAGSGGPGSGH 480
Db      421 TSGTTTSSGAASGKAGTGCTAGTTTSSBAGSDKAGTGTSGTTTSSGTGAGGAGSGGPGSGH 480
Qy      481 ASNAKIPGIMTLTLFALLTFIVN 503
Db      481 ASNAKIPGIMTLTLFALLTFIVN 503

RESULT 3
; Sequence 52, Application US/09737178
; Patent No. US20010029295A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Homer, Mary
; APPLICANT: Secretist, Heather
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C9
; CURRENT APPLICATION NUMBER: US/09/737,178
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 52
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-737-178-52

```

Query Match	100.0%	Score	2565;	DB	10;	Length	503;
Best Local Similarity	100.0%;	Prod. No.	9.3e-157;				
Matches	503;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	KRFEHDTMNGIHYYIDGSLLAGSEVTSNPRYSKEVEYEHTELAKEHCKKCKVCNVND	60				
Db	1	KRFEHDTMNGIHYYIDGSLLAGSEVTSNPRYSKEVEYEHTELAKEHCKKCKVCNVND	60				
Qy	61	IEDNNLKIYAKQFKSVVTPPADVAGVSDGPFIRQNLGAGVSNVEQPTVGMSLBQFIKN	120				
Db	61	IEDNNLKIYAKQFKSVVTPPADVAGVSDGPFIRQNLGAGVSNVEQPTVGMSLBQFIKN	120				
Qy	121	ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLRGA	180				
Db	121	ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLRGA	180				
Qy	181	RSEGNMFPQEALIRFRNASSEEMVNAASYLSAALFRYKEPFDEDLFKKANDNFRGDDGYDF	240				
Db	181	RSEGNMFPQEALIRFRNASSEEMVNAASYLSAALFRYKEPFDEDLFKKANDNFRGDDGYDF	240				
Qy	241	DYINTKKELVILASVLDGDLIMERLIENFSDVNTDDIKKAFDECKSNAIILKKKILDN	300				
Db	241	DYINTKKELVILASVLDGDLIMERLIENFSDVNTDDIKKAFDECKSNAIILKKKILDN	300				
Qy	301	DEDYKINPREMVNEVTCANTKFEALNDLIISDCEKKGIKINRDVISSYKLLSLTITYVG	360				
Db	301	DEDYKINPREMVNEVTCANTKFEALNDLIISDCEKKGIKINRDVISSYKLLSLTITYVG	360				
Qy	361	AGVEAVTVSVSATNGTBSGAGSGTGTSVSATSTLTGNGGTSGGTAGTTTSSGTBAGG	420				
Db	361	AGVEAVTVSVSATNGTBSGAGSGTGTSVSATSTLTGNGGTSGGTAGTTTSSGTBAGG	420				
Qy	421	TSGTTTSSGAASGKAGTGATGTTTSSGAGSDKAGTGTSGTTTSSGTCAGAGSGGPGSGH	480				

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:47:39 ; Search time 82.2782 Seconds  
(without alignments)  
726.026 Million cell updates/sec

Title: US-09-853-079-52

Perfect score: 2565

Sequence: 1 KRNEHTDMNGIHHYYIDGS.....AKIPGIMTLTLLFALLFFIVN 503

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2565	100.0	503	9	US-09-286-488-52
2	2565	100.0	503	9	US-09-853-079-52
3	2565	100.0	503	10	US-09-737-178-52
4	2239.5	87.3	452	9	US-09-853-079-209
5	2130	83.0	492	9	US-09-286-488-21
6	2130	83.0	492	9	US-09-853-079-21
7	2130	83.0	492	10	US-09-737-178-21
8	1159.5	45.2	1132	9	US-09-853-079-87
9	1159.5	45.2	1132	10	US-09-737-178-87
10	1143	44.6	666	9	US-09-853-079-85
11	1143	44.6	666	10	US-09-737-178-85
12	1141	44.5	677	9	US-09-853-079-144
13	1141	44.5	677	10	US-09-737-178-144
14	246.5	9.6	367	9	US-09-286-488-49
15	246.5	9.6	367	9	US-09-853-079-49
16	246.5	9.6	367	10	US-09-737-178-49
17	239.5	9.3	367	9	US-09-286-488-20
18	239.5	9.3	367	9	US-09-853-079-20
19	239.5	9.3	367	10	US-09-737-178-20

20	227.5	8.9	303	9	US-09-853-079-176	Sequence 176, App
21	227.5	8.9	309	9	US-09-853-079-121	Sequence 121, App
22	227.5	8.9	309	10	US-09-737-178-121	Sequence 121, App
23	222.5	8.7	309	9	US-09-853-079-136	Sequence 136, App
24	222.5	8.7	309	10	US-09-737-178-136	Sequence 136, App
25	219	8.5	45	9	US-09-853-079-208	Sequence 208, App
26	217.5	8.5	1395	9	US-10-123-155-3	Sequence 3, Appli
27	217.5	8.5	1395	9	US-10-146-731-3	Sequence 3, Appli
28	217	8.5	645	10	US-09-919-172-41	Sequence 41, Appl
29	212	8.3	592	9	US-09-286-488-24	Sequence 24, Appl
30	212	8.3	592	9	US-09-853-079-24	Sequence 24, Appl
31	212	8.3	592	10	US-09-737-178-24	Sequence 24, Appl
32	211	8.2	2478	9	US-10-184-644-107	Sequence 107, App
33	211	8.2	2478	9	US-10-184-634-107	Sequence 107, App
34	206.5	8.1	54	9	US-09-853-079-207	Sequence 207, App
35	206.5	8.1	2694	9	US-10-184-644-207	Sequence 207, App
36	206.5	8.1	2694	9	US-10-184-634-207	Sequence 207, App
37	202	7.9	1825	9	US-10-123-155-155	Sequence 155, App
38	202	7.9	1825	9	US-10-146-731-155	Sequence 155, App
39	199.5	7.8	1657	9	US-10-123-155-433	Sequence 433, App
40	199.5	7.8	1657	9	US-10-146-731-433	Sequence 433, App
41	199	7.8	3534	9	US-10-184-644-417	Sequence 417, App
42	199	7.8	3534	9	US-10-184-634-417	Sequence 417, App
43	198.5	7.7	3305	9	US-10-184-644-87	Sequence 87, Appl
44	198.5	7.7	3305	9	US-10-184-634-87	Sequence 87, Appl
45	198	7.7	1079	9	US-09-820-843A-20	Sequence 20, Appl

#### ALIGNMENTS

RESULT 1  
US-09-286-488-52  
; Sequence 52, Application US/09286488  
; Patent No. US20020169136A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C3  
; CURRENT APPLICATION NUMBER: US/09/286,488  
; CURRENT FILING DATE: 1999-04-05  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 52  
; LENGTH: 503  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-286-488-52

Query Match 100.0%; Score 2565; DB 9; Length 503;  
Best Local Similarity 100.0%; Pred No. 9,3e-157;  
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KRNEHTDMNGIHHYYIDGSLASGEVTSNFRYISKEYEYHEHTLAKEHCKEKCYNVDN 60  
Db 1 KRNEHTDMNGIHHYYIDGSLASGEVTSNFRYISKEYEYHEHTLAKEHCKEKCYNVDN 60  
QY 61 IEDNLLKIYAKFKSVVTTTPADVAGVSDGFFTRGQNLGAVGSVNEQPTVGSLEQFIKN 120  
Db 61 IEDNLLKIYAKFKSVVTTTPADVAGVSDGFFTRGQNLGAVGSVNEQPTVGSLEQFIKN 120  
QY 121 ELYSFSNEIYHTISSOISNSFLIMMSDALVKHDNYTLKKEGGCGEOIYNEEFIEKLRC 180  
Db 121 ELYSFSNEIYHTISSOISNSFLIMMSDALVKHDNYTLKKEGGCGEOIYNEEFIEKLRC 180  
QY 181 RSEGNMFMQALIRFRNASSEEMVNAASYSALFRYKFDDELFFKANDNFRDGDYDF 240  
Db 181 RSEGNMFMQALIRFRNASSEEMVNAASYSALFRYKFDDELFFKANDNFRDGDYDF 240

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF290205; AAL27891.1; JOINED.

DR EMBL; AF290206; AAL27891.1; JOINED.

DR EMBL; AF290207; AAL27891.1; JOINED.

SO SEQUENCE 465 AA; 49160 MW; 52A263D653DA8F80 CRC64;

Query Match 33.3%; Score 36; DB 11; Length 465;

Best Local Similarity 26.9%; Pred. No. 34;

Matches 7; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

Qy 6 NXNKSXXAXXKSDTQTXQEXXXE 31

Db 23 NGNETSAIKSTVQSHQSATTSTE 48

RESULT 13

Q9NXA5

ID Q9NXA5 PRELIMINARY; PRT; 692 AA.

AC Q9NXA5;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE CDNA FLJ20354 fis, clone HEP15013.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,

RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,

RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isegai T., Sugano S.,

RT "NEDO human cDNA sequencing project."

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK000361; BA91111.1;

DR InterPro; IPR000591; DEP.

DR SMART; SM00049; DEP; 1.

DR PROSITE; PS0186; DEP; 1.

SO SEQUENCE 692 AA; 79146 MW; 9B6101C81EFC8564 CRC64;

Query Match 33.3%; Score 36; DB 4; Length 692;

Best Local Similarity 33.3%; Pred. No. 52;

Matches 8; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

Qy 2 HXNKNKXAXXKSDTQTXQE 25

Db 470 HSENIQKPSAGFRTSLTVQD 493

RESULT 14

Q9H945

ID Q9H945 PRELIMINARY; PRT; 914 AA.

AC Q9H945;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE CDNA FLJ13022 fis, clone NT2RP3000753, weakly similar to neurofilament

DE TRIPLET H protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,

RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,

RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,

RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,

RA Masuho Y., Kanehori K.,

RT "NEDO human cDNA sequencing project."

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK023084; BAB14396.1;

SO SEQUENCE 914 AA; 100583 MW; 7E7A9B3A5BAA53C2 CRC64;

Query Match 33.3%; Score 36; DB 4; Length 914;

Best Local Similarity 44.4%; Pred. No. 69;

Matches 8; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 4 KXNKNKXAXXKSDTQ 21

Db 79 KAEDNQSIASVESGDTQ 96

RESULT 15

Q9AER7

ID Q9AER7 PRELIMINARY; PRT; 2402 AA.

AC Q9AER7;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Bap-like protein Bnp.

OS Staphylococcus epidermidis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Staphylococcus.

OX NCBI\_TaxID=1282;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RP62A;

RA Tormo M.A., Cucarella C., Amorena B., Lasa I., Penades J.R.;

RT "The Bap Homolog Protein of Staphylococcus epidermidis RP62A Promotes

RT Biofilm Formation."

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY028618; AAK29746.1;

DR InterPro; IPR001298; Filamin.

DR InterPro; IPR001899; Gram\_pos\_anchor.

DR InterPro; IPR003410; Hyalin.

DR InterPro; IPR000601; PKD\_domain.

DR Pfam; PF00746; Gram\_pos\_anchor; 1.

DR Pfam; PF02494; Hyr; 1.

DR Pfam; PF00801; PKD; 1.

DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.

DR TIGRFAMs; TIGR01168; YSIRK\_signal; 1.

DR PROSITE; PS0194; FILAMIN\_REPEAT; 2.

DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.

SO SEQUENCE 2402 AA; 258095 MW; D5807D96B8F2E9CC CRC64;

Query Match 33.3%; Score 36; DB 2; Length 2402;

Best Local Similarity 41.2%; Pred. No. 1.9e+02;

Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 16 KSXDTQTXQEXXXE 32

Db 125 KSADTQTTNETTNKND 141

Search completed: July 16, 2003, 17:50:01

Job time : 9.77066 secs



```

Query Match      34.3%; Score 37; DB 2; Length 241;
Best Local Similarity 36.4%; Pred. No. 11;
Matches      8; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

Oy 1 GHXXNKNKXXXXXXKSDTQT 22
    || | | | | | | | |
Db 30 GHPTRANRKSIVANKKNATQT 51

RESULT 7
ID O9KSW6 PRELIMINARY; PRT; 290 AA.
AC O9KSW6:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pseudouridine synthase family 1 protein.
GN VC1140.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=686;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004194; AAF94299.1; -
DR TIGR; VC1140; -
DR InterPro; IPR004397; Cons_hypoth93.
DR InterPro; IPR000613; Pseudod_synth.
DR InterPro; IPR000748; Psi_RSU.
DR Pfam; PF00849; Pseudod_synth.2; 1.
DR TIGRFAMs; TIGR00093; Cons_hypoth93; 1.
DR PROSITE; PS01149; PSI_RSU; 1.
KW Complete proteome.
SQ SEQUENCE 290 AA; 32611 MW; 283CC4062C205585 CRC64;

Query Match      34.3%; Score 37; DB 16; Length 290;
Best Local Similarity 36.4%; Pred. No. 13;
Matches      8; Conservative 1; Mismatches 13; Indels 0; Gaps 0;

Oy 1 GHXXNKNKXXXXXXKSDTQT 22
    || | | | | | | | |
Db 79 GHPTRANRKSIVANKKNATQT 100

RESULT 8
ID Q9LH95 PRELIMINARY; PRT; 608 AA.
AC Q9LH95:
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

```

```

[2]
RN SEQUENCE FROM N.A.
RP STRAIN=COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL; AP002057; BAB03177.1; -
DR InterPro; IPR002965; P_Rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 608 AA; 62766 MW; 6C6368AF4BD2A3B9 CRC64;

Query Match      34.3%; Score 37; DB 10; Length 608;
Best Local Similarity 29.2%; Pred. No. 28;
Matches      7; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

Oy 2 HXXNKNKXXXXXXKSDTQTQXE 25
    : | : | : | : | : | : |
Db 429 NSKSSSSSTTSVKVETQTSSE 452

RESULT 9
ID Q8SZ45 PRELIMINARY; PRT; 198 AA.
AC Q8SZ45:
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE REL7222p.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celisner S.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY071126; AAL48748.1; -
SQ SEQUENCE 198 AA; 22012 MW; CE33F831FAB79D49 CRC64;

Query Match      33.3%; Score 36; DB 5; Length 198;
Best Local Similarity 29.0%; Pred. No. 14;
Matches      9; Conservative 3; Mismatches 19; Indels 0; Gaps 0;

Oy 2 HXXNKNKXXXXXXKSDTQTQEXXXXXXEE 32
    | | | | | | | | | | | | | |
Db 115 HEFPNNEEQATRRSESTQATGPTQTE 145

RESULT 10
ID Q12288 PRELIMINARY; PRT; 251 AA.
AC Q12288:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ORF YLR126C.
GN YLR126C OR L3105 OR L3101 OR L9233.12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

```

```

DR pfam: PF00047; ig; 5.
DR pfam: PF00069; pkinaase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk pkinaase; 1.
DR SMART: SM00408; IGC2; 3.
DR SMART: SM00410; IG-like; 2.
DR SMART: SM00219; TYRK; 1.
DR SMART: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TVR; 1.
KW ATP-binding; Immunoglobulin domain; Transferase;
KW Tyrosine-protein kinase.
SQ SEQUENCE 1033 AA; 114266 MW; 559659540BDE66F5 CRC64;

Query Match 35.6%; Score 38.5; DB 5; Length 1033;
Best Local Similarity 40.0%; Pred. No. 24;
Matches 10; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

QY 1 GHKXNNKXXXXKXSDTQTXQE 25
DB 648 GHSK-SRSKSGDAQKSDDTACSQ 671

RESULT 5
O07532 PRELIMINARY; PRT; 488 AA.
AC 007532;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Endopeptidase LYTF precursor (Cell wall-associated polypeptide
DE CWP49).
GN LYTF OR YHDD.
OS Bacillus subtilis.
OC Bacteria; firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., SEQUENCE OF 27-37 AND CHARACTERIZATION.
RX MEDLINE=99140137; PubMed=10206711;
RA Margot P., Pagni M., Karamata D.;
RT "Bacillus subtilis 168 gene lytf encodes a gamma-D-glutamate-meso-
RT diaminopimelate muropetidase expressed by the alternative vegetative
RT sigma factor, sigma-D.";
RL Microbiology 145:57-65(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita Y., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Fujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

```

```

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzeneegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: D-GLUTAMATE - M-DIAMINOPIMELATE ENDOPEPTIDASE. CELL WALL
CC HYDROLASE INVOLVED IN CELL AUTOLYSIS.
CC -!- SUBCELLULAR LOCATION: CELL-WALL BOUND.
CC -!- SIMILARITY: BELONGS TO THE E.COLI NLPC / LISTERIA P60 FAMILY.
DR EMBL: Y14079; CAA74437.1; -.
DR EMBL: Z99109; CAB12776.1; -.
DR MEROPS: C40.002; -.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR000064; NLPC_P60.
DR Pfam: PF01476; LysM; 5.
DR Pfam: PF00877; NLPC_P60; 1.
DR SMART: SM00257; LysM; 5.
KW Cell wall; Hydrolase; Repeat; Signal; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 488 ENDOPEPTIDASE LYTF.
FT DOMAIN 140 146 POLY-SER.
FT DOMAIN 154 163 POLY-SER.
FT DOMAIN 166 172 POLY-SER.
FT DOMAIN 288 299 POLY-SER.
FT DOMAIN 305 310 POLY-THR.
FT DOMAIN 355 363 POLY-SER.
SQ SEQUENCE 488 AA; 51397 MW; 6FE0E96A4B42BDC8 CRC64;

Query Match 35.2%; Score 38; DB 16; Length 488;
Best Local Similarity 36.8%; Pred. No. 14;
Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 4 KXNNKXKXXKXSDTQT 22
DB 148 KSNKSSSSSSKSSSNKS 166

RESULT 6
O07532 PRELIMINARY; PRT; 241 AA.
AC 007532;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Pseudouridylyl synthase-like protein.
GN YMFC.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=566;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=569B;
RA Haralalka S., Roychoudhury S., Chaudhuri K.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF261151; AAG16128.1; -.
DR InterPro: IPR004397; Cons_hypoth93.
DR InterPro: IPR000613; Pseudou_synth.
DR InterPro: IPR000748; Psi_RSU.
DR Pfam: PF00849; Pseudou_synth_2; 1.
DR TIGRFAMs: TIGR00093; Cons_hypoth93; 1.
DR PROSITE: PS01149; PSI_RSU; 1.
SQ SEQUENCE 241 AA; 26964 MW; F12D39E9091079F CRC64;

```



DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE Seroreactive antigen BN1-20 precursor (Fragment).  
OS Babesia microti.  
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Babesiidae; Babesia.  
OX NCBI\_TaxID=5868;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MNI;  
RA Lodes M.J., Houghton R.L., Bruinsma E.S., Mohamath R., Reynolds L.D.,  
RT Benson D.R., Krause P.J., Reed S.G., Persing D.H.;  
RT "Serological expression cloning of novel immunoreactive antigens of  
RT Babesia microti";  
RL Infect. Immun. 68:2783-2790(2000).  
DR EMBL; AF206527; AAF68255.1; -  
KW SIGNAL.  
FT SIGNAL 1 22 POTENTIAL.  
FT NON\_TER 275 275  
FT SEQUENCE 275 AA; 30775 MW; C3CC5D5EBC494E01 CRC64;  
SQ  
Query Match 73.1%; Score 79; DB 5; Length 275;  
Best Local Similarity 56.2%; Pred. No. 2.4e-08;  
Matches 18; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 1 GHXXNKNKXXXXXKSDTQTXOEXXXXXEE 32  
DB 216 GHGKPTNKSEKAKRSHDTQTQTEICEECE 247  
RESULT 3  
Q24327  
ID Q24327 PRELIMINARY; PRT; 1033 AA.  
AC Q24327;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE GP160-DTRK precursor.  
GN OTK OR DTRK OR CG8967.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CANTON S;  
RA Pulido D., Campuzano S., Koda T., Modolell J., Barbacid M.;  
RT "Dtrk, a Drosophila gene related to the trk family of neurotrophin  
RT receptors, encodes a novel class of neural cell adhesion molecule";  
RL EMBO J. 11:391-404(1992).  
DR EMBL; X63453; CAA45053.1; -  
DR HSSP; P12931; 1FMK.  
DR FlyBase; FBgn0004839; otk.  
DR InterPro; IPR000719; Euk\_kinase.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR003600; Ig\_like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF00047; Ig\_5.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Euk\_kinase; 1.  
DR SMART; SM00408; IGC2; 3.  
DR SMART; SM00410; IG\_like; 2.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Cell adhesion; Immunoglobulin domain; Signal;  
KW transferase; Tyrosine-protein kinase.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 1033 GP160-DTRK.  
FT SEQUENCE 1033 AA; 114367 MW; D282EFCB28AC8D0 CRC64;

Query Match 35.6%; Score 38.5; DB 5; Length 1033;  
Best Local Similarity 40.0%; Pred. No. 24;  
Matches 10; Conservative 3; Mismatches 11; Indels 1; Gaps 1;  
QY 1 GHXXNKNKXXXXXKSDTQTXOE 25  
DB 648 GHSK-SRSGSGDAQKSDDTACSOO 671  
RESULT 4  
Q9V643  
ID Q9V643 PRELIMINARY; PRT; 1033 AA.  
AC Q9V643;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE OTK protein.  
GN OTK OR CG8967.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballev R.M., Basu P.V., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadenot E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dou L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Hougk J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003823; AAF58596.1; -  
DR HSSP; P12931; 1FMK.  
DR FlyBase; FBgn0004839; otk.  
DR InterPro; IPR000719; Euk\_kinase.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR003600; Ig\_like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR001245; Tyr\_kinase.

DT	01-OCT-2000 (TREMBlrel. 15, Created)
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)



		Query Match	9.6%; Score 246.5; DB 4; Length 367.
		Best Local Similarity	28.8%; Pred. No. 3.6e-11;
		Matches	92; Conservative 53; Mismatches 113; Indels 61; Gaps 12
Qy	80	PADVAGYSD--GFFIRGNLGAAGSVNQPTVNGSLSEQFIKNELYSFSPSNEIYHHISSOI	137
Dd	75	PSEAGGPSEAGPGSGTGSEAGWPSGTGWPSSEAGWSSERF--GYQLLPYSRRI--VFIFEV	131
Qy	138	SNSFLIMMSDAIVK-----HDNVI-----LKREGGECEQIYNVEEFIEKLGRAR	181
Dd	132	CLSIIYKHSMILERDRVDNDGHKDYLEBTKREKNKLKELEKC-----FPEQY----	179
Qy	182	SEGNNMFQEALIR-FRNASSEMVNAASYLSAALLFRYKFEDDELFKKANDFNRDGDGYDF	240
Dd	180	--SLMKKEELARIFDNAST-----ISS---KYLLLVDEISNKAYGLTEGPAADNF	224
Qy	241	DYINTKKELVILASVLOGDLIMIERLIENPFDVNNTDDIKAFDECKSNAILKKKILDN	300
Dd	225	DHFNIWKSIVLKDMFTCYDLLLOHLIYKFYYDYNTVDNIKFNDESKALVLRDKITKK	284
Qy	301	DEDYKINFPMRVNEVTCAHTKFEALNDLIISDCEKKGIKINRDVISSYKKLLISTTYI--	358
Dd	285	DGDYNTHPEDMIKELNSAEENFKVIDMISNI----GDYDEYDSIASEKPFISMITEYTK	341
Qy	359	-----VGAGEAVETVSV	370
Dd	342	ITKVSNVIIPIGKALTITV	360

RESULT 14  
US-08-845-258-20  
; Sequence 20, Application US/08845258  
; Patent No. 6183976  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,258
; FILING DATE: 24-APR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206)682-6031
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 367 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-845-258-20
Query Match 9.3%; Score 239.5; DB 4 Length 367

```

STATE: Washington  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/845,258  
FILING DATE: 24-APR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.426C1  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 367 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-845-258-49

Query Match 9.6%; Score 246.5; DB 4; Length 367;  
Best Local Similarity 28.8%; Pred. No. 3.6e-11;  
Matches 92; Conservative 53; Mismatches 113; Indels 61; Gaps 12;  
QY 80 PADVAGVSD--GFFIRGQNLGAVGSVNEQNTVGMSELEQFIKNELYSFSNEIYHTISSQI 137  
Db 75 PSEAGGSEAGGPGSGTGSEAGWPSGTGWSEAGWSSERF--GYQLLPYSRRI--VIFNEV 131  
QY 138 SNSEFLIMSDAIYK-----HDNYI-----LKEGEGCEQIYNYEEFIEKURGAR 181  
Db 132 CLSYIYKHVSMLERDRVNDGHDKYIEETKEKNKLKLEK-----FPEQY---- 179  
QY 182 SEGNNFQEOALIR--FRNASSEEMVNAASLYSALFRYKFEDELFKANDNFGDDGYDF 240  
Db 180 ---SLMKKEELARIFDNAST-----ISS---KYLLVDEISNKAYTGLEGPADNF 224  
QY 241 DYINTKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 300  
Db 225 DFRNINWKSIVLKDIFYCDLLQHLIYKFYDNTVNDIKKNFDESKKALVLRDKITTK 284  
QY 301 DEDYKINFRWNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYI-- 358  
Db 285 DGYNTHFEDMIKELNSAAEFNKIVDIMISNI---G DYDEYDSIASFKPFLSMITEITK 341  
QY 359 -----VGAGVEAVTVSV 370  
Db 342 ITKVSNVPIPGIKALTLTV 360

## RESULT 11

US-08-990-571-49

Sequence 49, Application US/08990571

Patent No. 6214971

GENERAL INFORMATION:

APPLICANT: Reed, Steven G. et al.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

NUMBER OF SEQUENCES: 79

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/990,571  
FILING DATE: 11-DEC-1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.426C2  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 367 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-990-571-49

Query Match 9.6%; Score 246.5; DB 4; Length 367;  
Best Local Similarity 28.8%; Pred. No. 3.6e-11;  
Matches 92; Conservative 53; Mismatches 113; Indels 61; Gaps 12;  
QY 80 PADVAGVSD--GFFIRGQNLGAVGSVNEQNTVGMSELEQFIKNELYSFSNEIYHTISSQI 137  
Db 75 PSEAGGSEAGGPGSGTGSEAGWPSGTGWSEAGWSSERF--GYQLLPYSRRI--VIFNEV 131  
QY 138 SNSEFLIMSDAIYK-----HDNYI-----LKEGEGCEQIYNYEEFIEKURGAR 181  
Db 132 CLSYIYKHVSMLERDRVNDGHDKYIEETKEKNKLKLEK-----FPEQY---- 179  
QY 182 SEGNNFQEOALIR--FRNASSEEMVNAASLYSALFRYKFEDELFKANDNFGDDGYDF 240  
Db 180 ---SLMKKEELARIFDNAST-----ISS---KYLLVDEISNKAYTGLEGPADNF 224  
QY 241 DYINTKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 300  
Db 225 DFRNINWKSIVLKDIFYCDLLQHLIYKFYDNTVNDIKKNFDESKKALVLRDKITTK 284  
QY 301 DEDYKINFRWNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYI-- 358  
Db 285 DGYNTHFEDMIKELNSAAEFNKIVDIMISNI---G DYDEYDSIASFKPFLSMITEITK 341  
QY 359 -----VGAGVEAVTVSV 370  
Db 342 ITKVSNVPIPGIKALTLTV 360

## RESULT 12

US-08-723-142A-49

Sequence 49, Application US/08723142A

Patent No. 6306396

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Lodes, Michael J.

APPLICANT: Houghton, Raymond

APPLICANT: Sleath, Paul R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

QY 61 IEDNNLKIIYAKQKSVVTTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSEQFIKN 120  
DB 133 IEDNNLKIIYAKQKSVVTTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSEQFIKN 192  
QY 121 ELYFSNEIYHTISSQISNFIIMMSDAIVKHDNYILKKEGCEQIYNEEFIEKLRGA 180  
DB 193 ELYFSNEIYHTISSQISNFIIMMSDAIVKHDNYILKKEGCEQIYNEEFIEKLRGA 252  
QY 181 RSEGNMFQALIRFNASSEEMVNAASVLSAALRYKEFDDDELKPKANDNFRDGDYDF 240  
DB 253 RSEGNMFQALIRFNASSEEMVNAASVLSAALRYKEFDDDELKPKANDNFRDGDYDF 312  
QY 241 DYINTKKELVILASVLDGLDLIERLIENFSDVNTDDIKKAFDECKSNAIILKKKILDN 300  
DB 313 DYINTKKELVILASVLDGLDLIERLIENFSDVNTDDIKKAFDECKSNAIILKKKILDN 372  
QY 301 DEDYKINPREMNEVTCANTKEALNDLIISDCEKKGKIKINRDVISSYKLLSTIYIYG 360  
DB 373 DEDYKINPREMNEVTCANTKEALNDLIISDCEKKGKIKINRDVISSYKLLSTIYIYG 432  
QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSILTNGGTESGGTAGTTSSGTGAG 419  
DB 433 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSILTNGGTESGGTAGTTSSGTGAG 491

RESULT 8  
US-09-528-784A-87  
; Sequence 87, Application US/09528784A  
; Patent No. 6451315  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C4  
; CURRENT APPLICATION NUMBER: US/09/528,784A  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 87  
; LENGTH: 1132  
; TYPE: PRT  
; ORGANISM: Babesia  
US-09-528-784A-87

Query Match 45.2%; Score 1159.5; DB 4; Length 1132;  
Best Local Similarity 58.3%; Pred. No. 3.2e-80;  
Matches 278; Conservative 35; Mismatches 89; Indels 75; Gaps 14;

QY 59 DNI-----EDNNLKIIYA-KQFKSVVTTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGM 112  
DB 269 DNIITTRNEVTKDDYALKKALTCTLTHLIVHSKVDG--ISPDMLGTQKNSSPLGKIGT 326  
QY 113 SLEQFIKNELYSFSEIYHTISSQISNFIIMMSD-AIVKHDNYILKKEGCEQIYNE 171  
DB 327 SMDIIFA-----MFSN-----PNMYLVKAVYLOAIEHIFLISTKYND-----IFDYT 368  
QY 172 -----EFTEKLRGARSEGNMNFQALIRFNASSEEMVNAASVLSAALRYKE 219  
DB 369 IDFSKREATDSGFTDILLGNKVRKESLFI EGLSIDKSHLAKAGVTGIGSSSLF----- 424  
QY 220 FDELFKCANDNFRDGDG-----YFDYINTKK-----ELVILASVLDGLDLIERL 266  
DB 425 --DEIF-----DELNDQATIRTLVAPLDWPLISDKSLHPSLKMVY---VLPGFIV--- 471  
QY 267 IENFSDVNTDDIKKAFDECKSNAIILKKKILDNDEYKINPREMNEVTCANTKEALN 326  
DB 472 -----PGSTDDIKKAFDECKSNAIILKKKILDNDEYKINPREMNEVTCANTKEALN 525

QY 327 DLIISDCEKKGKIKINRDVISSYKLLSTIYIYGAGVEAVTVSVSATSNGTSGGAGSGT 386  
DB 526 DLIISDCEKKGKIKINRDVISSYKLLSTIYIYGAGVEAVTVSVSATSNGTSGGAGSGT 585  
QY 387 GTSVSTATSLTNGGTESGGTAGTTSSGTERAGGTSGTTSSGAAAGKAGTGTAGTTTSS 446  
DB 586 GTSVSTATSLTNGGTESGGTAGTTSSGTERAGGTSGTTSSGAAAGKAGTGTAGTTTSS 645  
QY 447 EGAGSDKAGTGTSGTTSSGTEAGGAGSGGSPGSHASNAKIPGIMTLTLFALLTFIVN 503  
DB 646 EGAGSDKAGTGTSGTTSSGTEAGGAGSGGSPGSHASNAKIPGIMTLTLFALLTFIVN 702

RESULT 9  
US-09-528-784A-85  
; Sequence 85, Application US/09528784A  
; Patent No. 6451315  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C4  
; CURRENT APPLICATION NUMBER: US/09/528,784A  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 85  
; LENGTH: 666  
; TYPE: PRT  
; ORGANISM: Babesia  
US-09-528-784A-85

Query Match 44.6%; Score 1143; DB 4; Length 666;  
Best Local Similarity 99.1%; Pred. No. 2.7e-79;  
Matches 228; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 274 NNTDDIKKAFDECKSNAIILKKKILDNDEYKINPREMNEVTCANTKEALNDLIISDC 333  
DB 7 HHTDDIKKAFDECKSNAIILKKKILDNDEYKINPREMNEVTCANTKEALNDLIISDC 66  
QY 334 EKKGKIKINRDVISSYKLLSTIYIYGAGVEAVTVSVSATSNGTSGGAGSGTGTSVSAT 393  
DB 67 EKKGKIKINRDVISSYKLLSTIYIYGAGVEAVTVSVSATSNGTSGGAGSGTGTSVSAT 126  
QY 394 SLTNGGTESGGTAGTTSSGTEAGGTSGTTSSGAAAGKAGTGTAGTTTSSGAGSDK 453  
DB 127 SLTNGGTESGGTAGTTSSGTEAGGTSGTTSSGAAAGKAGTGTAGTTTSSGAGSDK 186  
QY 454 AGTGTSGTTSSGTEAGGAGSGGSPGSHASNAKIPGIMTLTLFALLTFIVN 503  
DB 187 AGTGTSGTTSSGTEAGGAGSGGSPGSHASNAKIPGIMTLTLFALLTFIVN 236

RESULT 10  
US-08-845-258-49  
; Sequence 49, Application US/08845258  
; Patent No. 6183976  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 492 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-990-571-21

Query Match 83.0%; Score 2130; DB 4; Length 492;  
Best Local Similarity 99.5%; Pred. No. 7.1e-155;  
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKEHCKKCVNVDN 60  
DB 73 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKEHCKKCVNVDN 132  
QY 61 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGONLGAVGSVNEQNTVGMSELEQFIKN 120  
DB 133 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGONLGAVGSVNEQNTVGMSELEQFIKN 192  
QY 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFTEKL RGA 180  
DB 193 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFTEKL RGA 252  
QY 181 RSEGNMFQALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKNDFGRDDGYDF 240  
DB 253 RSEGNMFQALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKNDFGRDDGYDF 312  
QY 241 DYINTKKELVLASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 300  
DB 313 DYINTKKELVLASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 372  
QY 301 DEDYKINFREMNVEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVG 360  
DB 373 DEDYKINFREMNVEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVG 432  
QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVTSATSLTNGGTSGGTAGTTSSGTWFG 419  
DB 433 AGVEAVTVSVSATSNGTSGGAGSGTGTSVTSATSLTNGGTSGGTAGTTSSGTWFG 491

## RESULT 6

US-08-723-142A-21  
; Sequence 21, Application US/08723142A  
; Patent No. 6306396  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/723,142A  
; FILING DATE: 01-OCT-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.426  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 492 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-723-142A-21

Query Match 83.0%; Score 2130; DB 4; Length 492;  
Best Local Similarity 99.5%; Pred. No. 7.1e-155;  
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKEHCKKCVNVDN 60  
DB 73 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKEHCKKCVNVDN 132  
QY 61 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGONLGAVGSVNEQNTVGMSELEQFIKN 120  
DB 133 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGONLGAVGSVNEQNTVGMSELEQFIKN 192  
QY 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFTEKL RGA 180  
DB 193 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFTEKL RGA 252  
QY 181 RSEGNMFQALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKNDFGRDDGYDF 240  
DB 253 RSEGNMFQALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKNDFGRDDGYDF 312  
QY 241 DYINTKKELVLASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 300  
DB 313 DYINTKKELVLASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 372  
QY 301 DEDYKINFREMNVEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVG 360  
DB 373 DEDYKINFREMNVEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVG 432  
QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVTSATSLTNGGTSGGTAGTTSSGTWFG 419  
DB 433 AGVEAVTVSVSATSNGTSGGAGSGTGTSVTSATSLTNGGTSGGTAGTTSSGTWFG 491

## RESULT 7

US-09-528-784A-21  
; Sequence 21, Application US/09528784A  
; Patent No. 6451315  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Sleath, Paul R.  
; APPLICANT: McNeill, Patricia D.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION  
; FILE REFERENCE: 210121.426C4  
; CURRENT APPLICATION NUMBER: US/09/528,784A  
; CURRENT FILING DATE: 2000-03-17  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 21  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Babesia microti  
US-09-528-784A-21

Query Match 83.0%; Score 2130; DB 4; Length 492;  
Best Local Similarity 99.5%; Pred. No. 7.1e-155;  
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKEHCKKCVNVDN 60  
DB 73 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKEHCKKCVNVDN 132

QY 1 KRFNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHCKKCKCVNVDN 60  
Db 1 KRFNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHCKKCKCVNVDN 60  
QY 61 IEDNNLKIIYAKQKSVVTPPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSLQFQIKN 120  
Db 61 IEDNNLKIIYAKQKSVVTPPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSLQFQIKN 120  
QY 121 ELYSFSNEIYHTISSQISNSFLIMSDAIVKHNDYILKKEGEGCEQIYNYEEFIEKLGA 180  
Db 121 ELYSFSNEIYHTISSQISNSFLIMSDAIVKHNDYILKKEGEGCEQIYNYEEFIEKLGA 180  
QY 181 RSEGNMFQALIRFNASSEEMVNAASYLSAALFRYKEFDDELFFKANDNFRDGDYDF 240  
Db 181 RSEGNMFQALIRFNASSEEMVNAASYLSAALFRYKEFDDELFFKANDNFRDGDYDF 240  
QY 241 DYINTKELVILASVLDGLDLIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 300  
Db 241 DYINTKELVILASVLDGLDLIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 300  
QY 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKINRDVISSYKLLSTITYIVG 360  
Db 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKINRDVISSYKLLSTITYIVG 360  
QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSTLTGNGGTESGAGTGTSSGTEAGG 420  
Db 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSTLTGNGGTESGAGTGTSSGTEAGG 420  
QY 421 TSGTTSSGAAGKAGTGAGTTTSSGAGSDKAGTGTTSSGAGGAGGAGGPGSH 480  
Db 421 TSGTTSSGAAGKAGTGAGTTTSSGAGSDKAGTGTTSSGAGGAGGAGGPGSH 480  
QY 481 ASNAKIPGIMTLTFLALLFIYN 503  
Db 481 ASNAKIPGIMTLTFLALLFIYN 503

RESULT 4

US-08-845-258-21

; Sequence 21, Application US/08845258

; Patent No. 6183976

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond

; APPLICANT: Sleath, Paul R.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; AND TREATMENT OF B. MICROTI INFECTION

; NUMBER OF SEQUENCES: 53

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; City: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/845,258

; FILING DATE: 24-APR-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.426C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206)682-6031

; INFORMATION FOR SEQ ID NO: 21:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 492 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; US-08-845-258-21

Query Match 83.0%; Score 2130; DB 4; Length 492;

Best Local Similarity 99.5%; Pred.No.7.1e-155;

Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KRFNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHCKKCKCVNVDN 60  
Db 73 KRFNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHCKKCKCVNVDN 132  
QY 61 IEDNNLKIIYAKQKSVVTPPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSLQFQIKN 120  
Db 133 IEDNNLKIIYAKQKSVVTPPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSLQFQIKN 192  
QY 121 ELYSFSNEIYHTISSQISNSFLIMSDAIVKHNDYILKKEGEGCEQIYNYEEFIEKLGA 180  
Db 193 ELYSFSNEIYHTISSQISNSFLIMSDAIVKHNDYILKKEGEGCEQIYNYEEFIEKLGA 252  
QY 181 RSEGNMFQALIRFNASSEEMVNAASYLSAALFRYKEFDDELFFKANDNFRDGDYDF 240  
Db 253 RSEGNMFQALIRFNASSEEMVNAASYLSAALFRYKEFDDELFFKANDNFRDGDYDF 312  
QY 241 DYINTKELVILASVLDGLDLIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 300  
Db 313 DYINTKELVILASVLDGLDLIMERLIENFSDVNTDDIKKAFDECKSNAILKKKILDN 372  
QY 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKINRDVISSYKLLSTITYIVG 360  
Db 373 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKINRDVISSYKLLSTITYIVG 432  
QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSTLTGNGGTESGAGTGTSSGTEAG 419  
Db 433 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSTLTGNGGTESGAGTGTSSGTEAG 491

RESULT 5

US-08-990-571-21

; Sequence 21, Application US/08990571

; Patent No. 6214971

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G. et al.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS

; AND TREATMENT OF B. MICROTI INFECTION

; NUMBER OF SEQUENCES: 79

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; City: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/990,571

; FILING DATE: 11-DEC-1997

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.426C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206)682-6031

; INFORMATION FOR SEQ ID NO: 21:



```
Qy 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKHECKKCKCVNVDN 60
Db 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKHECKKCKCVNVDN 60
Qy 61 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMLEQFIKN 120
Db 61 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMLEQFIKN 120
Qy 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHNDYILKKEGEGCEQIYNYEEFIEKLGA 180
Db 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHNDYILKKEGEGCEQIYNYEEFIEKLGA 180
Qy 181 RSEGNMFOBALIRFRNASSEEMVNAASYLSAALFRYKEFDDELKANDNFRDGDGYDF 240
Db 181 RSEGNMFOBALIRFRNASSEEMVNAASYLSAALFRYKEFDDELKANDNFRDGDGYDF 240
Qy 241 DYINTKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300
Db 241 DYINTKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300
Qy 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVVISSYKLLSTIYIVG 360
Db 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVVISSYKLLSTIYIVG 360
Qy 361 AGVEAVTVSVSATSNGTESGAGSGTGTSVSATSTLTGNGGTSAGTGTSSGTEAGG 420
Db 361 AGVEAVTVSVSATSNGTESGAGSGTGTSVSATSTLTGNGGTSAGTGTSSGTEAGG 420
Qy 421 TSGTTSSGAGSKAGTGAGTTSSRGAGSDKAGTGTSVTSATSTLTGNGGTSAGTGTSSG 480
Db 421 TSGTTSSGAGSKAGTGAGTTSSRGAGSDKAGTGTSVTSATSTLTGNGGTSAGTGTSSG 480
Qy 481 ASNAKIPGIMTLTLFALLTFIVN 503
Db 481 ASNAKIPGIMTLTLFALLTFIVN 503

RESULT 2
us-09-990-571-52
; Sequence 52, Application US/08990571
; Patent No. 6214971
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B. M
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,571
; FILING DATE: 11-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Babesia Microti
; US-08-990-571-52

Query Match 100.0%; Score 2565; DB 4; Length 503;
Best Local Similarity 100.0%; Pred. No. 4.3e-188;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKHECKKCKCVNVDN 60
Db 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSISKEYEYEHTELAKHECKKCKCVNVDN 60
Qy 61 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMLEQFIKN 120
Db 61 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMLEQFIKN 120
Qy 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHNDYILKKEGEGCEQIYNYEEFIEKLGA 180
Db 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHNDYILKKEGEGCEQIYNYEEFIEKLGA 180
Qy 181 RSEGNMFOBALIRFRNASSEEMVNAASYLSAALFRYKEFDDELKANDNFRDGDGYDF 240
Db 181 RSEGNMFOBALIRFRNASSEEMVNAASYLSAALFRYKEFDDELKANDNFRDGDGYDF 240
Qy 241 DYINTKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300
Db 241 DYINTKELVILASVLDGLDIMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300
Qy 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVVISSYKLLSTIYIVG 360
Db 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVVISSYKLLSTIYIVG 360
Qy 361 AGVEAVTVSVSATSNGTESGAGSGTGTSVSATSTLTGNGGTSAGTGTSSGTEAGG 420
Db 361 AGVEAVTVSVSATSNGTESGAGSGTGTSVSATSTLTGNGGTSAGTGTSSGTEAGG 420
Qy 421 TSGTTSSGAGSKAGTGAGTTSSRGAGSDKAGTGTSVTSATSTLTGNGGTSAGTGTSSG 480
Db 421 TSGTTSSGAGSKAGTGAGTTSSRGAGSDKAGTGTSVTSATSTLTGNGGTSAGTGTSSG 480
Qy 481 ASNAKIPGIMTLTLFALLTFIVN 503
Db 481 ASNAKIPGIMTLTLFALLTFIVN 503

RESULT 3
US-09-528-784A-52
; Sequence 52, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C4
; CURRENT APPLICATION NUMBER: US/09/528,784A
; CURRENT FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 52
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Babesia microti
; US-09-528-784A-52

Query Match 100.0%; Score 2565; DB 4; Length 503;
Best Local Similarity 100.0%; Pred. No. 4.3e-188;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:42:54 ; Search time 45.8044 Seconds  
(without alignments)  
323.107 Million cell updates/sec

Title: US-09-853-079-52

Perfect score: 2565

Sequence: 1 KRFEHDTMGTHYHYIDGS.....AKIPGIMTLTFLALITFIVN 503

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2565	100.0	503	US-08-845-258-52	Sequence 52, Appl
2	2565	100.0	503	US-08-990-571-52	Sequence 52, Appl
3	2565	100.0	503	US-09-528-784A-52	Sequence 52, Appl
4	2130	83.0	492	US-08-845-258-21	Sequence 21, Appl
5	2130	83.0	492	US-08-990-571-21	Sequence 21, Appl
6	2130	83.0	492	US-08-723-142A-21	Sequence 21, Appl
7	2130	83.0	492	US-09-528-784A-21	Sequence 21, Appl
8	1159.5	45.2	1132	US-09-528-784A-87	Sequence 87, Appl
9	1143	44.6	666	US-09-528-784A-85	Sequence 85, Appl
10	246.5	9.6	367	US-08-845-258-49	Sequence 49, Appl
11	246.5	9.6	367	US-08-990-571-49	Sequence 49, Appl
12	246.5	9.6	367	US-08-723-142A-49	Sequence 49, Appl
13	246.5	9.6	367	US-09-528-784A-49	Sequence 49, Appl
14	239.5	9.3	367	US-08-845-258-20	Sequence 20, Appl
15	239.5	9.3	367	US-08-990-571-20	Sequence 20, Appl
16	239.5	9.3	367	US-08-723-142A-20	Sequence 20, Appl
17	239.5	9.3	367	US-09-528-784A-20	Sequence 20, Appl
18	212	8.3	592	US-08-845-258-24	Sequence 24, Appl
19	212	8.3	592	US-08-990-571-24	Sequence 24, Appl
20	212	8.3	592	US-08-723-142A-24	Sequence 24, Appl
21	212	8.3	592	US-09-528-784A-24	Sequence 24, Appl
22	194.5	7.6	1038	US-07-609-716-36	Sequence 36, Appl
23	194.5	7.6	1038	US-08-475-411A-36	Sequence 36, Appl
24	194.5	7.6	1038	US-08-478-029A-36	Sequence 36, Appl
25	192	7.5	1011	US-08-477-509B-94	Sequence 94, Appl
26	192	7.5	1011	US-08-482-085B-94	Sequence 94, Appl
27	192	7.5	1011	US-09-444-791A-94	Sequence 94, Appl

28 192 7.5 1170 1 US-08-175-155-59 Sequence 59, Appl  
29 192 7.5 1170 2 US-08-707-237A-66 Sequence 66, Appl  
30 191 7.4 463 4 US-08-845-258-25 Sequence 25, Appl  
31 191 7.4 463 4 US-08-990-571-25 Sequence 25, Appl  
32 191 7.4 463 4 US-08-723-142A-25 Sequence 25, Appl  
33 191 7.4 463 4 US-09-528-784A-25 Sequence 25, Appl  
34 189.5 7.4 1177 1 US-07-609-716-31 Sequence 31, Appl  
35 189.5 7.4 1177 1 US-08-175-155-29 Sequence 29, Appl  
36 189.5 7.4 1177 1 US-08-477-509B-64 Sequence 64, Appl  
37 189.5 7.4 1177 2 US-08-707-237A-35 Sequence 35, Appl  
38 189.5 7.4 1177 3 US-08-482-085B-64 Sequence 64, Appl  
39 189.5 7.4 1177 4 US-08-475-411A-31 Sequence 31, Appl  
40 189.5 7.4 1177 4 US-08-478-029A-31 Sequence 31, Appl  
41 189.5 7.4 1177 4 US-09-444-791A-64 Sequence 64, Appl  
42 188.5 7.3 1059 1 US-08-175-155-48 Sequence 48, Appl  
43 188.5 7.3 1059 2 US-08-707-237A-54 Sequence 54, Appl  
44 188.5 7.3 1059 4 US-08-806-029-10 Sequence 10, Appl  
45 188.5 7.3 1101 1 US-08-477-509B-83 Sequence 83, Appl

#### ALIGNMENTS

#### RESULT 1

US-08-845-258-52  
; Sequence 52, Application US/08845258  
; Patent No. 6183976  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Sleath, Paul R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS  
; TITLE OF INVENTION: AND TREATMENT OF B. MICROID INFECTION  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: SEED AND BERRY  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/845,258  
; FILING DATE: 24-APR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.426C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206)682-6031  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 503 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Babesia Microti  
US-08-845-258-52

Query Match 100.0%; Score 2565; DB 4; Length 503;  
Best Local Similarity 100.0%; Pred. No. 4.3e-188;  
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



CC coding sequences. The B. microti antigens, antigenic epitopes of such  
CC antigens, and compositions comprising such antigens are useful for  
CC diagnosing and treating B. microti infection. The compositions are  
CC especially useful for enhancing immune response against B. microti  
CC infection. The present sequence was used to illustrate the invention.

XX  
SQ Sequence 666 AA;  
Query Match 44.6%; Score 1143; DB 23; Length 666;  
Best Local Similarity 99.1%; Pred. No. 8.9e-69;  
Matches 228; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 274 NNTDDIKAFDECKSNAILKKILNDNEDYKINFRMNVNEVTCANTKFEALNDLIISDC 333  
DB 7 HHTDDIKAFDECKSNAILKKILNDNEDYKINFRMNVNEVTCANTKFEALNDLIISDC 66  
QY 334 EKKGKINRDVVISSYKLLSTIYIVGAGVEAVTVSVATSNGTSGGAGSGTGSVSAT 393  
DB 67 EKKGKINRDVVISSYKLLSTIYIVGAGVEAVTVSVATSNGTSGGAGSGTGSVSAT 126  
QY 394 STLTNGGTSGGTAGTTSSGTAGTSGTSSGAAAGKAGTGTAGTTSSGAGSDK 453  
DB 127 STLTNGGTSGGTAGTTSSGTAGTSGTSSGAAAGKAGTGTAGTTSSGAGSDK 186  
QY 454 AGTGTSGTTSSGTGAGGAGGPGSGHASNAPKIPGIMTLTLFALLTFIVN 503  
DB 187 AGTGTSGTTSSGTGAGGAGGPGSGHASNAPKIPGIMTLTLFALLTFIVN 236

RESULT 14  
ABB88989  
ID ABB88989 standard; Protein; 677 AA.  
AC ABB88989;  
XX  
XX  
DT 20-JUN-2002 (first entry)  
XX  
DE Babesia microti antigenic epitope fusion protein Baf-5.  
KW Protozoacide; vaccine; antigen; antigenic epitope; infection.  
XX Babesia microti.  
OS  
XX  
PN WO200185947-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 09-MAY-2001; 2001WO-US15192.  
XX  
PR 10-MAY-2000; 2000US-0569098.  
PR 27-JUN-2000; 2000US-0605724.  
PR 07-SEP-2000; 2000US-0656688.  
PR 10-OCT-2000; 2000US-0685436.  
PR 13-DEC-2000; 2000US-0737178.  
PR 26-FEB-2001; 2001US-0794764.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;  
PI Secrist H;  
XX  
XX WPI; 2002-216691/27.  
XX  
XX New Babesia microti antigens, useful for diagnosing and treating B.  
PT microti infection, and as component of a composition for enhancing  
PT immune response against B. microti infections  
XX  
XX Claim 35; Page 160-163; 195pp; English.  
XX  
XX The present invention relates to novel Babesia microti antigens and their  
CC coding sequences. The B. microti antigens, antigenic epitopes of such  
CC antigens, and compositions comprising such antigens are useful for  
CC diagnosing and treating B. microti infection. The compositions are

CC especially useful for enhancing immune response against B. microti.  
CC infection. The present sequence was used to illustrate the invention.

XX  
SQ Sequence 677 AA;  
Query Match 44.5%; Score 1141; DB 23; Length 677;  
Best Local Similarity 100.0%; Pred. No. 1.2e-68;  
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 TDDIKAFDECKSNAILKKILNDNEDYKINFRMNVNEVTCANTKFEALNDLIISDC 335  
DB 450 TDDIKAFDECKSNAILKKILNDNEDYKINFRMNVNEVTCANTKFEALNDLIISDC 509  
QY 336 KGKINRDVVISSYKLLSTIYIVGAGVEAVTVSVATSNGTSGGAGSGTGSVSATST 395  
DB 510 KGKINRDVVISSYKLLSTIYIVGAGVEAVTVSVATSNGTSGGAGSGTGSVSATST 569  
QY 396 LTGNGGTSGGTAGTTSSGTAGTSGTSSGAAAGKAGTGTAGTTSSGAGSDK 455  
DB 570 LTGNGGTSGGTAGTTSSGTAGTSGTSSGAAAGKAGTGTAGTTSSGAGSDK 629  
QY 456 TGTSGTTSSGTGAGGAGGPGSGHASNAPKIPGIMTLTLFALLTFIVN 503  
DB 630 TGTSGTTSSGTGAGGAGGPGSGHASNAPKIPGIMTLTLFALLTFIVN 677

RESULT 15  
AAW56301  
ID AAW56301 standard; Protein; 367 AA.  
AC AAW56301;  
XX  
XX  
DT 28-SEP-1998 (first entry)  
XX  
DE Babesia microti antigen sequence.  
KW antigen; detection; diagnosis; vaccine; tick-borne disease;  
KW differentiation; Lyme disease; ehrlichiosis.  
XX Babesia microti.  
OS  
XX  
PN EP834567-A2.  
XX  
PD 08-APR-1998.  
XX  
PF 01-OCT-1997; 97EP-0117067.  
XX  
PR 24-APR-1997; 97US-0845258.  
PR 01-OCT-1996; 96US-0723142.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Houghton R, Lodes MJ, Reed SG, Sleath PR;  
XX WPI; 1998-195465/18.  
XX  
XX Polypeptides comprising Babesia microti antigens and their  
PT immunogenic fragments or epitopes - and related nucleic acid,  
PT vectors, transformed cells and antibodies, useful for diagnosis of  
PT infection and in protective vaccines  
XX  
XX Disclosure; Page 94-95; 113pp; English.  
XX  
XX The sequence is that of a polypeptide comprising at least  
CC one antigenic portion of a Babesia microti antigen. It can be used  
CC to diagnose B. microti infection by detecting specific antibodies  
CC in usual immunoassays. Infection can also be diagnosed using:  
CC (a) primers or probes derived from the coding sequence, in  
CC standard amplification or hybridisation tests, or (b) using  
CC antibodies to detect the corresponding antigen. It is also  
CC useful in vaccines to protect against infection, especially  
CC when formulated with an adjuvant. The new diagnostic methods  
CC allow rapid differentiation between B. microti infection and



QY 241 DYINTKKELVILASVLDGLDLIMERLIENFSDVNNNTDDIKKAFDECKSNAILKKKILDN 300  
 DB 313 DYINTKKELVILASVLDGLDLIMERLIENFSDVNNNTDDIKKAFDECKSNAILKKKILDN 372  
 QY 301 DEDYKINREMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVLISSYKLLLSITITYIVG 360  
 DB 373 DEDYKINREMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVLISSYKLLLSITITYIVG 432  
 QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSFLTNGGTGSGTAGTTSSGTEAG 419  
 DB 433 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSFLTNGGTGSGTAGTTSSGTEAG 491  
 RESULT 10  
 AAB30231  
 ID AAB30231 standard; Protein; 1132 AA.  
 XX AAB30231;  
 XX 12-FEB-2001 (first entry)  
 DT B. microti MN-10/BMNI-17/BMNI-15 fusion protein SEQ ID NO: 87.  
 DE Babesiosis; rodent parasite; tick-borne illness; antigen;  
 KW disease diagnosis; disease prevention.  
 XX Babesia sp.  
 OS Synthetic.  
 OS WO200060090-A1.  
 PN 12-OCT-2000.  
 PD 05-APR-2000; 2000WO-US09136.  
 PF 05-APR-1999; 99US-0286488.  
 PR 17-MAR-2000; 2000US-0528784.  
 XX (CORI-) CORIXA CORP.  
 XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;  
 XX WPI; 2000-686939/67.  
 XX New polypeptides containing an antigenic portion of Babesia microti  
 PT antigen and DNAs encoding the polypeptides, useful for diagnosing,  
 PT treating or preventing B. microti infection, or for inducing protective  
 PT immunity in a patient -  
 XX Example 7; Page 112-116; 118pp; English.  
 XX The present invention is related to the isolation of antigenic sequences  
 CC from the rodent parasite Babesia microti. This organism is transmitted to  
 CC humans by the same tick which transmits Lyme disease and ehrlichiosis.  
 CC The organism causes a malaria-like infection known as babesiosis. The  
 CC sequences identified by this invention can be used in the diagnosis,  
 CC prevention and treatment of babesiosis.  
 XX Sequence 1132 AA;  
 SQ Query Match 45.2%; Score 1159.5; DB 21; Length 1132;  
 Best Local Similarity 58.3%; Pred. No. 1.4e-69;  
 Matches 278; Conservative 35; Mismatches 89; Indels 75; Gaps 14;  
 QY 59 DNI-----EDNNKIYA-KQKSVVTPADVAGSDGFFIRGQNLGAVSGVNEQPNVGM 112  
 DB 269 DNIITTRNEVTKDDVYALKKALCTLTTHLYHSKVDG--ISFDMLGTQKNKSSPLGKIGT 326  
 QY 113 SLEQFIKNELYSFSENIYHTISSQISNLSFLIMSD-AIVKHDNVLKKEGEGCEQIYNYE 171  
 DB 327 SMDDIIA-----MFSN-----PNNYLVKVAYLQAIHFIFLSTKYND-----IFDIT 368  
 QY 172 -----EFIEKLRGARSEGNMNFQALIFRNASSEMVNAASVLSAALFRYKE 219

DB 369 IDFSKREATDSGFTDILLGNKVKESLSFIEGLISDIKSHSLKAGVTGGISSSLF----- 424  
 QY 220 FDDDLFFKANDNFRDDG-----YDFDYINTKK-----ELVILASVLDGLDLIMERL 266  
 DB 425 --DEIF-----DELNDQATIRTLVAPLDWPLDISKLSLHPSLKMVV---VLPGFVIV---- 471  
 QY 267 IENFSDVNNNTDDIKKAFDECKSNAILKKKILDNDYKINFRMNEVTCANTKFEALN 326  
 DB 472 -----PGSTDDIKKAFDECKSNAILKKKILDNDYKINFRMNEVTCANTKFEALN 525  
 QY 327 DLIISDCEKKGKIKINRDVLISSYKLLLSITITYIVGAGVEAVTVSVSATSNGTSGGAGSGT 386  
 DB 526 DLIISDCEKKGKIKINRDVLISSYKLLLSITITYIVGAGVEAVTVSVSATSNGTSGGAGSGT 585  
 QY 387 GTSVSATSLTNGGTGSGTAGTTSSGTEAGTSGTSSGAASGKAGTGAGTTTSS 446  
 DB 586 GTSVSATSLTNGGTGSGTAGTTSSGTEAGTSGTSSGAASGKAGTGAGTTTSS 645  
 QY 447 EGAGSDKAGTGTCTTTSSGTGAGGAGSGGPGSHASNAKIPGIMTLFLFALLFIIVN 503  
 DB 646 EGAGSDKAGTGTCTTTSSGTGAGGAGSGGPGSHASNAKIPGIMTLFLFALLFIIVN 702  
 RESULT 11  
 ABB88976  
 ID ABB88976 standard; Protein; 1132 AA.  
 XX ABB88976;  
 XX 20-JUN-2002 (first entry)  
 DT Babesia microti antigenic epitope fusion protein BaP-3.  
 DE Protozoacide; vaccine; antigen; antigenic epitope; infection.  
 KW Babesia microti.  
 OS WO200185947-A2.  
 PN 15-NOV-2001.  
 PD 09-MAY-2001; 2001WO-US15192.  
 PF 10-MAY-2000; 2000US-0569098.  
 PR 27-JUN-2000; 2000US-0605724.  
 PR 07-SEP-2000; 2000US-0656688.  
 PR 10-OCT-2000; 2000US-0685436.  
 PR 13-DEC-2000; 2000US-0737178.  
 PR 26-FEB-2001; 2001US-0794764.  
 XX (CORI-) CORIXA CORP.  
 XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;  
 XX Secrist H;  
 XX WPI; 2002-216691/27.  
 XX New Babesia microti antigens, useful for diagnosing and treating B.  
 PT microti infection, and as component of a composition for enhancing  
 PT immune response against B. microti infections -  
 XX Claim 35; Page 116-120; 195pp; English.  
 XX The present invention relates to novel Babesia microti antigens and their  
 CC coding sequences. The B. microti antigens, antigenic epitopes of such  
 CC antigens, and compositions comprising such antigens are useful for  
 CC diagnosing and treating B. microti infection. The compositions are  
 CC especially useful for enhancing immune response against B. microti  
 CC infection. The present sequence was used to illustrate the invention.  
 XX Sequence 1132 AA;  
 SQ

AC AAB30191;  
XX 12-FEB-2001 (first entry)  
XX B. microti BMNI-4 antigen SEQ ID NO: 21.  
XX Babesiosis; rodent parasite; tick-borne illness; antigen;  
XX disease diagnosis; disease prevention.  
XX Babesia microti.  
XX WO200060090-A1.  
XX 12-OCT-2000.  
XX 05-APR-2000; 2000WO-US09136.  
XX 05-APR-1999; 99US-0286488.  
XX 17-MAR-2000; 2000US-0528784.  
XX (CORI-) CORIXA CORP.  
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;  
XX WPI: 2000-686939/67.  
XX N-PSDB: AAC65081.  
XX New polypeptides containing an antigenic portion of Babesia microti  
XX antigen and DNAs encoding the polypeptides, useful for diagnosing,  
XX treating or preventing B. microti infection, or for inducing protective  
XX immunity in a patient  
XX Disclosure; Page 72-73; 118pp; English.  
XX  
XX The present invention is related to the isolation of antigenic sequences  
XX from the rodent parasite Babesia microti. This organism is transmitted to  
XX humans by the same tick which transmits Lyme disease and ehrlichiosis.  
XX The organism causes a malaria-like infection known as babesiosis. The  
XX sequences identified by this invention can be used in the diagnosis,  
XX prevention and treatment of babesiosis.  
XX  
XX Sequence 492 AA;  
XX  
XX Query Match 83.0%; Score 2130; DB 21; Length 492;  
XX Best Local Similarity 99.5%; Pred. No. 1.9e-135;  
XX Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKEHCKKCVNVDN 60  
Db 73 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKEHCKKCVNVDN 132  
Qy 61 IEDNNLIYAKQKSVVTTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSEQFIKN 120  
Db 133 IEDNNLIYAKQKSVVTTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSEQFIKN 192  
Qy 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNVEEFTKURGA 180  
Db 193 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNVEEFTKURGA 252  
Qy 181 RSEGNMFQALIRFRNASSEEMVNAAYLSAALFRYKEFDELFKKNDFGRDGDYDF 240  
Db 253 RSEGNMFQALIRFRNASSEEMVNAAYLSAALFRYKEFDELFKKNDFGRDGDYDF 312  
Qy 241 DYINTKRELVLASVLDGLDILMERLIENFSDVNTDDIKAFDECKSNAILLKKILDN 300  
Db 313 DYINTKRELVLASVLDGLDILMERLIENFSDVNTDDIKAFDECKSNAILLKKILDN 372  
Qy 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINROVISSYKLLSTITVIG 360  
Db 373 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINROVISSYKLLSTITVIG 432  
Qy 361 AGVEAVTVSVSATSNGTSGAGSGTGTSVSAATSLTNGCGTSGGTAGTTTSSGTWFG 491

Db 433 AGVEAVTVSVSATSNGTSGAGSGTGTSVSAATSLTNGCGTSGGTAGTTTSSGTWFG 491  
RESULT 9  
ABB88937  
ID ABB88937 standard; Protein; 492 AA.  
XX ABB88937;  
XX 20-JUN-2002 (first entry)  
XX Babesia microti antigen BMNI-4.  
XX Protozoacide; vaccine; antigen; antigenic epitope; infection.  
XX Babesia microti.  
XX WO200185947-A2.  
XX 15-NOV-2001.  
XX 09-MAY-2001; 2001WO-US15192.  
XX 10-MAY-2000; 2000US-0569098.  
XX 27-JUN-2000; 2000US-0605724.  
XX 07-SEP-2000; 2000US-0656688.  
XX 10-OCT-2000; 2000US-0685436.  
XX 13-DEC-2000; 2000US-0737178.  
XX 26-FEB-2001; 2001US-0794764.  
XX (CORI-) CORIXA CORP.  
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;  
XX Secrist H;  
XX WPI: 2002-216691/27.  
XX N-PSDB: ABL89325.  
XX New Babesia microti antigens, useful for diagnosing and treating B.  
XX microti infection, and as component of a composition for enhancing  
XX immune response against B. microti infections  
XX  
XX Example 1; Page 78-79; 195pp; English.  
XX  
XX The present invention relates to novel Babesia microti antigens and their  
XX coding sequences. The B. microti antigens, antigenic epitopes of such  
XX antigens, and compositions comprising such antigens are useful for  
XX diagnosing and treating B. microti infection. The compositions are  
XX especially useful for enhancing immune response against B. microti  
XX infection. The present sequence was used to illustrate the invention.  
XX  
XX Sequence 492 AA;  
XX  
XX Query Match 83.0%; Score 2130; DB 23; Length 492;  
XX Best Local Similarity 99.5%; Pred. No. 1.9e-135;  
XX Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKEHCKKCVNVDN 60  
Db 73 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKEHCKKCVNVDN 132  
Qy 61 IEDNNLIYAKQKSVVTTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSEQFIKN 120  
Db 133 IEDNNLIYAKQKSVVTTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSEQFIKN 192  
Qy 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNVEEFTKURGA 180  
Db 193 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNVEEFTKURGA 252  
Qy 181 RSEGNMFQALIRFRNASSEEMVNAAYLSAALFRYKEFDELFKKNDFGRDGDYDF 240  
Db 253 RSEGNMFQALIRFRNASSEEMVNAAYLSAALFRYKEFDELFKKNDFGRDGDYDF 312

XX (CORI-) CORIXA CORP.  
XX Houghton R, Lodes MJ, Reed SG, Sleath PR;  
XX WPI; 1998-195465/18.  
DR N-PSDB; AAV22734.  
XX  
XX Polypeptides comprising Babesia microti antigens and their  
PT immunogenic fragments or epitopes - and related nucleic acid.  
PT vectors, transformed cells and antibodies, useful for diagnosis of  
PT infection and in protective vaccines  
XX  
XX Claim 1; Page 49-51; 113pp; English.  
PS  
XX The sequence is that of a polypeptide comprising at least  
CC one antigenic portion of a Babesia microti antigen. It can be used  
CC to diagnose B. microti infection by detecting specific antibodies  
CC in usual immunoassays. Infection can also be diagnosed using:  
CC (a) primers or probes derived from the coding sequence, in  
CC standard amplification or hybridisation tests, or (b) using  
CC antibodies to detect the corresponding antigen. It is also  
CC useful in vaccines to protect against infection, especially  
CC when formulated with an adjuvant. The new diagnostic methods  
CC allow rapid differentiation between B. microti infection and  
CC other tick-borne diseases (Lyme disease and ehrlichiosis) that  
CC have similar symptoms but require different treatments.  
XX  
XX Sequence 492 AA;  
SQ  
Query Match 83.0%; Score 2130; DB 19; Length 492;  
Best Local Similarity 99.5%; Pred. No. 1.9e-135;  
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KRNEHTDMNGTHYYIDGSLASGEVTSNFRYISKEYEYETELAKEHCKKEKCVNDN 60  
DB 73 KRNEHTDMNGTHYYIDGSLASGEVTSNFRYISKEYEYETELAKEHCKKEKCVNDN 132  
QY 61 IEDNNLKIIYAKQFKSVVTPADVAGVSDGFFIRGQNLGAVGVNEQPTVGMSEOFIKN 120  
DB 133 IEDNNLKIIYAKQFKSVVTPADVAGVSDGFFIRGQNLGAVGVNEQPTVGMSEOFIKN 192  
QY 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLGA 180  
DB 193 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLGA 252  
QY 181 RSEGNMFQALIRFNASSEEMVNAASYLSAALFRYKFEFDDDELFFKANDNFRDGDYDF 240  
DB 253 RSEGNMFQALIRFNASSEEMVNAASYLSAALFRYKFEFDDDELFFKANDNFRDGDYDF 312  
QY 241 DYINTKKELVILASVLDGLDLIMERLIENFSDVNTDDIKKAFDECKSNAIILKKKILDN 300  
DB 313 DYINTKKELVILASVLDGLDLIMERLIENFSDVNTDDIKKAFDECKSNAIILKKKILDN 372  
QY 301 DEDYKINFREMYNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVG 360  
DB 373 DEDYKINFREMYNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVG 432  
QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSSTLTNGGTESGAGTGTSSGTEAG 419  
DB 433 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSSTLTNGGTESGAGTGTSSGTEAG 491  
RESULT 7  
AAY24342  
ID AAY24342 standard; Protein; 492 AA.  
XX  
AC AAY24342;  
XX  
DT 16-SEP-1999 (first entry)  
XX  
DE Babesia microti antigen BMNI-4.  
XX

KW Babesia microti; antigen; immunogen; diagnosis; infection; vaccine;  
KW immunity; detection.  
XX Babesia microti.  
OS  
XX  
PN WO9929869-A1.  
XX  
XX 17-JUN-1999.  
XX  
XX 11-DEC-1998; 98WO-US26437.  
XX  
XX 11-DEC-1997; 97US-0990571.  
XX  
XX (CORI-) CORIXA CORP.  
PA (MAYO-) MAYO FOUNDATION.  
XX  
XX Bruisma E, Houghton R, Lodes MJ, Persing D, Reed SG;  
PI Sleath PR;  
XX  
XX WPI; 1999-385612/32.  
DR N-PSDB; AAX88998.  
XX  
XX New isolated Babesia microti polypeptides  
PT  
XX  
XX Example 1; Page 72-74; 126pp; English.  
XX  
XX The present invention describes isolated polypeptides comprising  
CC specific immunogenic portions of Babesia microti. AAX88983 to AAX88994  
CC encode specifically claimed B. microti immunogenic proteins, and  
CC AAY24327 to AAY24338 represent these proteins. B. microti polypeptides  
CC and nucleic acids can be used for detecting B. microti infections. They  
CC can also be used in vaccines for inducing protective immunity against B.  
CC microti infections. The present sequence represents a B. microti antigen.  
XX  
SQ Sequence 492 AA;  
Query Match 83.0%; Score 2130; DB 20; Length 492;  
Best Local Similarity 99.5%; Pred. No. 1.9e-135;  
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KRNEHTDMNGTHYYIDGSLASGEVTSNFRYISKEYEYETELAKEHCKKEKCVNDN 60  
DB 73 KRNEHTDMNGTHYYIDGSLASGEVTSNFRYISKEYEYETELAKEHCKKEKCVNDN 132  
QY 61 IEDNNLKIIYAKQFKSVVTPADVAGVSDGFFIRGQNLGAVGVNEQPTVGMSEOFIKN 120  
DB 133 IEDNNLKIIYAKQFKSVVTPADVAGVSDGFFIRGQNLGAVGVNEQPTVGMSEOFIKN 192  
QY 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLGA 180  
DB 193 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNYEEFIEKLGA 252  
QY 181 RSEGNMFQALIRFNASSEEMVNAASYLSAALFRYKFEFDDDELFFKANDNFRDGDYDF 240  
DB 253 RSEGNMFQALIRFNASSEEMVNAASYLSAALFRYKFEFDDDELFFKANDNFRDGDYDF 312  
QY 241 DYINTKKELVILASVLDGLDLIMERLIENFSDVNTDDIKKAFDECKSNAIILKKKILDN 300  
DB 313 DYINTKKELVILASVLDGLDLIMERLIENFSDVNTDDIKKAFDECKSNAIILKKKILDN 372  
QY 301 DEDYKINFREMYNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVG 360  
DB 373 DEDYKINFREMYNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITYIVG 432  
QY 361 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSSTLTNGGTESGAGTGTSSGTEAG 419  
DB 433 AGVEAVTVSVSATSNGTSGGAGSGTGTSVSATSSTLTNGGTESGAGTGTSSGTEAG 491  
RESULT 8  
AAB30191  
ID AAB30191 standard; Protein; 492 AA.  
XX



QY 61 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSSLEQFIK 120  
 DB 61 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSSLEQFIK 120  
 QY 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNVEEFTEKLRGA 180  
 DB 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNVEEFTEKLRGA 180  
 QY 181 RSEGNMFQEQALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKNANDNFRDGDYDF 240  
 DB 181 RSEGNMFQEQALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKNANDNFRDGDYDF 240  
 QY 241 DYINTKKELVILASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300  
 DB 241 DYINTKKELVILASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300  
 QY 301 DEDYKINFREMNVEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLLSITIVIG 360  
 DB 301 DEDYKINFREMNVEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLLSITIVIG 360  
 QY 361 AGVEAVTVSVSATSNGTSSGAGSGTGTSTLTGNGGTESGGTAGTTSSGTEAGG 420  
 DB 361 AGVEAVTVSVSATSNGTSSGAGSGTGTSTLTGNGGTESGGTAGTTSSGTEAGG 420  
 QY 421 TSGTTSSGAASGKAGTGAGTTSSGAGSDKAGTGTSTLTGNGGTESGGTAGTTSSGTEAGG 480  
 DB 421 TSGTTSSGAASGKAGTGAGTTSSGAGSDKAGTGTSTLTGNGGTESGGTAGTTSSGTEAGG 480  
 QY 481 ASNAKIPGIMTLTLFALLTFIVN 503  
 DB 481 ASNAKIPGIMTLTLFALLTFIVN 503

## RESULT 5

ABB89024  
 ID ABB89024 standard; Protein; 452 AA.  
 XX  
 AC ABB89024;  
 XX  
 DT 20-JUN-2002 (first entry)  
 XX  
 DE Babesia microti antigen MN-10 variant.  
 XX  
 KW Protozoacide; vaccine; antigen; antigenic epitope; infection.  
 XX  
 OS Babesia microti.  
 XX  
 PN WO200185947-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 09-MAY-2001; 2001WO-US15192.  
 XX  
 PR 10-MAY-2000; 2000US-0569098.  
 PR 27-JUN-2000; 2000US-0605724.  
 PR 07-SEP-2000; 2000US-0656688.  
 PR 10-OCT-2000; 2000US-0685436.  
 PR 13-DEC-2000; 2000US-0737178.  
 PR 26-FEB-2001; 2001US-0794764.  
 XX  
 PA (CORI-) CORIAX CORP.  
 XX  
 PI Reed SG, Lodes MJ, Houghton RL, McNeill PR, McNeill PD, Homer MJ;  
 PI Secrist H;  
 XX  
 DR WPI; 2002-216691/27.  
 XX  
 PT New Babesia microti antigens, useful for diagnosing and treating B.  
 PT microti infection, and as component of a composition for enhancing  
 PT immune response against B. microti infections  
 XX  
 PS Claim 2; Page 188-190; 195pp; English.  
 XX

CC The present invention relates to novel Babesia microti antigens and their  
 CC coding sequences. The B. microti antigens, antigenic epitopes of such  
 CC antigens, and compositions comprising such antigens are useful for  
 CC diagnosing and treating B. microti infection. The compositions are  
 CC especially useful for enhancing immune response against B. microti  
 CC infection. The present sequence was used to illustrate the invention.  
 XX  
 SQ Sequence 452 AA;  
 Query Match 87.3%; Score 2239.5; DB 23; Length 452;  
 Best Local Similarity 89.0%; Pred. No. 7.3e-143;  
 Matches 445; Conservative 0; Mismatches 4; Indels 51; Gaps 1;  
 QY 1 KRNEHTDMGHIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHECKKCKVCNVND 60  
 DB 4 KRNEHTDMGHIHYYIDGSLASGEVTSNFRYISKEYEYEHTELAKHECKKCKVCNVND 63  
 QY 61 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSSLEQFIK 120  
 DB 64 IEDNNLIYAKQKFSVVTTPADVAGVSDGFFIRGQNLGAVGSVNEQNTVGMSSLEQFIK 123  
 QY 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNVEEFTEKLRGA 180  
 DB 124 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNVEEFTEKLRGA 183  
 QY 181 RSEGNMFQEQALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKNANDNFRDGDYDF 240  
 DB 184 RSEGNMFQEQALIRFRNASSEEMVNAASYLSAALFRYKEFDELFKKNANDNFRDGDYDF 243  
 QY 241 DYINTKKELVILASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300  
 DB 244 DYINTKKELVILASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 303  
 QY 301 DEDYKINFREMNVEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLLSITIVIG 360  
 DB 304 DEDYKINFREMNVEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLLSITIVIG 363  
 QY 361 AGVEAVTVSVSATSNGTSSGAGSGTGTSTLTGNGGTESGGTAGTTSSGTEAGG 420  
 DB 364 AGVEAVTVSVSATSNGTSSGAGSGTGTSTLTGNGGTESGGTAGTTSSGTEAGG 420  
 QY 421 TSGTTSSGAASGKAGTGAGTTSSGAGSDKAGTGTSTLTGNGGTESGGTAGTTSSGTEAGG 480  
 DB 393 -----GTTSSGAGSGKAGTGAGTTSSGAGSDKAGTGTSTLTGNGGTESGGTAGTTSSGTEAGG 432  
 QY 481 ASNAKIPGIMTLTLFALLTF 500  
 DB 433 ASNAKIPGIMTLTLFALLTF 452  
 RESULT 6  
 AAW56281  
 ID AAW56281 standard; Protein; 492 AA.  
 XX  
 AC AAW56281;  
 XX  
 DT 28-SEP-1998 (first entry)  
 XX  
 DE Babesia microti BMNI-4 antigen sequence.  
 XX  
 KW antigen; detection; diagnosis; vaccine; tick-borne disease;  
 KW differentiation; Lyme disease; ehrlichiosis.  
 XX  
 OS Babesia microti.  
 XX  
 PN EP834567-A2.  
 XX  
 PD 08-APR-1998.  
 XX  
 PF 01-OCT-1997; 97EP-0117067.  
 XX  
 PR 24-APR-1997; 97US-0845258.  
 PR 01-OCT-1996; 96US-0723142.

```
RESULT 3
AAB30216
ID AAB30216 standard; Protein; 503 AA.
XX
AC AAB30216;
XX
DT 12-FEB-2001 (first entry)
XX
DE B. microti clone antigen SEQ ID NO: 52.
XX
KW Babesiosis; rodent parasite; tick-borne illness; antigen;
KW disease diagnosis; disease prevention.
XX
OS Babesia microti.
XX
PN WO200060090-A1.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-US09136.
XX
PR 05-APR-1999; 99US-0286488.
PR 17-MAR-2000; 2000US-0528784.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;
XX
DR WPI; 2000-686939/67.
XX
PT New polypeptides containing an antigenic portion of Babesia microti
PT antigen and DNAs encoding the polypeptides, useful for diagnosing,
PT treating or preventing B. microti infection, or for inducing protective
PT immunity in a patient
XX
PS Claim 14; Page 96-98; 118pp; English.
XX
CC The present invention is related to the isolation of antigenic sequences
CC from the rodent parasite Babesia microti. This organism is transmitted to
CC humans by the same tick which transmits Lyme disease and ehrlichiosis.
CC The organism causes a malaria-like infection known as babesiosis. The
CC sequences identified by this invention can be used in the diagnosis,
CC prevention and treatment of babesiosis.
XX
SQ Sequence 503 AA;

Query Match 100.0%; Score 2565; DB 21; Length 503;
Best Local Similarity 100.0%; Pred. No. 9.8e-165;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KRFEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYHEHTELAKEHCKKCVNVDN 60
DB 1 KRFEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYHEHTELAKEHCKKCVNVDN 60
QY 61 IEDNNLKIIYAKQFQKSVVTPADVAGVSDGFFIRGQNLGAVGSVNEQPTVGNLSLEQFIKN 120
DB 61 IEDNNLKIIYAKQFQKSVVTPADVAGVSDGFFIRGQNLGAVGSVNEQPTVGNLSLEQFIKN 120
QY 121 ELYSFSNEYHTIHTSSQISNFIIMMSDALVKHDNIIKKEGRCQIYNYEFIEKLRGA 180
DB 121 ELYSFSNEYHTIHTSSQISNFIIMMSDALVKHDNIIKKEGRCQIYNYEFIEKLRGA 180
QY 181 RSEGNMFQEALIRFNASSEMVNAA SYLSAALFRYKEFDDDELFKKANDNFRDGDGYDF 240
DB 181 RSEGNMFQEALIRFNASSEMVNAA SYLSAALFRYKEFDDDELFKKANDNFRDGDGYDF 240
QY 241 DYINTKKELVILASVLDGLDLIMERLIENFSDVNNNTDDIKKAFDECKSNAIILKKIILDN 300
DB 241 DYINTKKELVILASVLDGLDLIMERLIENFSDVNNNTDDIKKAFDECKSNAIILKKIILDN 300
QY 301 DEDYKINFREVMNEVTCANTKFEALNDLIISDCEKKGIKINRDVISSYKLLLTITYIVG 360

DB 301 DEDYKINFREVMNEVTCANTKFEALNDLIISDCEKKGIKINRDVISSYKLLLTITYIVG 360
QY 361 AGVEAVTVSVSATSNCTESGGAGSGTGTSVSATSLTNGGTTESGCTAGTTTSSGTEAGG 420
DB 361 AGVEAVTVSVSATSNCTESGGAGSGTGTSVSATSLTNGGTTESGCTAGTTTSSGTEAGG 420
QY 421 TSGTTTSSGAASGKAGTGAGTTTSSGAGSDKAGTGTSCTTTSSGTGAGGAGSGGPSGH 480
DB 421 TSGTTTSSGAASGKAGTGAGTTTSSGAGSDKAGTGTSCTTTSSGTGAGGAGSGGPSGH 480
QY 481 ASNAKIPGIMTLTLFALLTFIVN 503
DB 481 ASNAKIPGIMTLTLFALLTFIVN 503

RESULT 4
AAB88961
ID AAB88961 standard; Protein; 503 AA.
XX
AC AAB88961;
XX
DT 20-JUN-2002 (first entry)
XX
DE Babesia microti antigen MN-10.
XX
KW Protozoacide; vaccine; antigen; antigenic epitope; infection.
XX
OS Babesia microti.
XX
PN WO200185947-A2.
XX
PD 15-NOV-2001.
XX
PF 09-MAY-2001; 2001WO-US15192.
XX
PR 10-MAY-2000; 2000US-0569098.
PR 27-JUN-2000; 2000US-0605724.
PR 07-SEP-2000; 2000US-0656688.
PR 10-OCT-2000; 2000US-0685436.
PR 13-DEC-2000; 2000US-0737178.
PR 26-FEB-2001; 2001US-0794764.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;
PI Secrlist H;
XX
WPI; 2002-216691/27.
XX
PT New Babesia microti antigens, useful for diagnosing and treating B.
PT microti infection, and as component of a composition for enhancing
PT immune response against B. microti infections
XX
PS Claim 14; Page 102-103; 195pp; English.
XX
CC The present invention relates to novel Babesia microti antigens and their
CC coding sequences. The B. microti antigens, antigenic epitopes of such
CC antigens, and compositions comprising such antigens are useful for
CC diagnosing and treating B. microti infection. The compositions are
CC especially useful for enhancing immune response against B. microti
CC infection. The present sequence was used to illustrate the invention.
XX
SQ Sequence 503 AA;

Query Match 100.0%; Score 2565; DB 23; Length 503;
Best Local Similarity 100.0%; Pred. No. 9.8e-165;
Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KRFEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYHEHTELAKEHCKKCVNVDN 60
DB 1 KRFEHTDMNGIHYYIDGSLASGEVTSNFRYISKEYEYHEHTELAKEHCKKCVNVDN 60
```

PT vectors, transformed cells and antibodies, useful for diagnosis of  
 XX infection and in protective vaccines

PS Claim 1; Page 99-101; 113pp; English.

XX The sequence is that of a polypeptide comprising at least  
 CC one antigenic portion of a Babesia microti antigen. It can be used  
 CC to diagnose B. microti infection by detecting specific antibodies  
 CC in usual immunoassays. Infection can also be diagnosed using:  
 CC (a) primers or probes derived from the coding sequence, in  
 CC standard amplification or hybridisation tests, or (b) using  
 CC antibodies to detect the corresponding antigen. It is also  
 CC useful in vaccines to protect against infection, especially  
 CC when formulated with an adjuvant. The new diagnostic methods  
 CC allow rapid differentiation between B. microti infection and  
 CC other tick-borne diseases (Lyme disease and ehrlichiosis) that  
 CC have similar symptoms but require different treatments.

XX Sequence 503 AA;

Query Match 100.0%; Score 2565; DB 19; Length 503;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-165;  
 Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSKYEYEHTELAKEHCKKCVNVDN 60  
 DB 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSKYEYEHTELAKEHCKKCVNVDN 60  
 QY 61 IEDNNLIYAKQKFSVVTTPADVAGSDGFFIRQNLGAVGSVNEQNTVGMSSLEQFIKN 120  
 DB 61 IEDNNLIYAKQKFSVVTTPADVAGSDGFFIRQNLGAVGSVNEQNTVGMSSLEQFIKN 120  
 QY 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNTYEEFTEKLURGA 180  
 DB 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNTYEEFTEKLURGA 180  
 QY 181 RSEGNMFQEQALIRFRNASSEEMVNAAYLSAALFRYKEFDELFKANDNFGDDCYDF 240  
 DB 181 RSEGNMFQEQALIRFRNASSEEMVNAAYLSAALFRYKEFDELFKANDNFGDDCYDF 240  
 QY 241 DYINTKRELVLASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300  
 DB 241 DYINTKRELVLASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300  
 QY 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITIVG 360  
 DB 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITIVG 360  
 QY 361 AGVEAVTVSVATSNGTESGAGSGTGTSTLTGNGTSGGTAGTTSSGTGAGG 420  
 DB 361 AGVEAVTVSVATSNGTESGAGSGTGTSTLTGNGTSGGTAGTTSSGTGAGG 420  
 QY 421 TSGTTTSSGAASGAGTGAGTTTSSGAGSDKAGTGTSTLTGNGTSGGTAGTTSSGTGAGG 480  
 DB 421 TSGTTTSSGAASGAGTGAGTTTSSGAGSDKAGTGTSTLTGNGTSGGTAGTTSSGTGAGG 480  
 QY 481 ASNAKIPGIMTLTLFALLTFIVN 503  
 DB 481 ASNAKIPGIMTLTLFALLTFIVN 503

RESULT 2

AY24364  
 ID AY24364 standard; Protein; 503 AA.

XX AC AY24364;

XX DT 16-SEP-1999 (first entry)

XX Babesia microti antigen MN-10 complementary open reading frame protein.

XX Babesia microti; antigen; immunogen; diagnosis; infection; vaccine;

KW immunity; detection.

XX Babesia microti.  
 XX OS WO9929869-A1.  
 XX PN 17-JUN-1999.  
 XX PD 11-DEC-1998; 98WO-US26437.  
 XX PF 11-DEC-1997; 97US-0990571.  
 XX PR (CORI-) CORIXA CORP.  
 XX PA (MAYO-) MAYO FOUNDATION.  
 XX PI Bruisma E, Houghton R, Lodes MJ, Persing D, Reed SG;  
 XX PI Sleath PR;  
 XX DR WPI; 1999-385612/32.  
 XX DR N-PSDB; AAX90016.  
 XX PT New Isolated Babesia microti polypeptides  
 XX PS Example 1; Page 107-108; 126pp; English.  
 XX CC The present invention describes isolated polypeptides comprising  
 CC specific immunogenic portions of Babesia microti. AAX88983 to AAX88994  
 CC encode specifically claimed B. microti immunogenic proteins, and  
 CC AAY24327 to AAY24338 represent these proteins. B. microti polypeptides  
 CC and nucleic acids can be used for detecting B. microti infections. They  
 CC can also be used in vaccines for inducing protective immunity against B.  
 CC microti infections. The present sequence represents a B. microti antigen.  
 XX SQ Sequence 503 AA;  
 Query Match 100.0%; Score 2565; DB 20; Length 503;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-165;  
 Matches 503; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSKYEYEHTELAKEHCKKCVNVDN 60  
 DB 1 KRNEHTDMNGIHYYIDGSLASGEVTSNFRYSKYEYEHTELAKEHCKKCVNVDN 60  
 QY 61 IEDNNLIYAKQKFSVVTTPADVAGSDGFFIRQNLGAVGSVNEQNTVGMSSLEQFIKN 120  
 DB 61 IEDNNLIYAKQKFSVVTTPADVAGSDGFFIRQNLGAVGSVNEQNTVGMSSLEQFIKN 120  
 QY 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNTYEEFTEKLURGA 180  
 DB 121 ELYSFSNEIYHTISSQISNSFLIMMSDAIVKHDNYILKKEGEGCEQIYNTYEEFTEKLURGA 180  
 QY 181 RSEGNMFQEQALIRFRNASSEEMVNAAYLSAALFRYKEFDELFKANDNFGDDCYDF 240  
 DB 181 RSEGNMFQEQALIRFRNASSEEMVNAAYLSAALFRYKEFDELFKANDNFGDDCYDF 240  
 QY 241 DYINTKRELVLASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300  
 DB 241 DYINTKRELVLASVLDGLDILMERLIENFSDVNTDDIKKAFDECKSNAILLKKKILDN 300  
 QY 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITIVG 360  
 DB 301 DEDYKINFRMNEVTCANTKFEALNDLIISDCEKKGKIKINRDVISSYKLLSTITIVG 360  
 QY 361 AGVEAVTVSVATSNGTESGAGSGTGTSTLTGNGTSGGTAGTTSSGTGAGG 420  
 DB 361 AGVEAVTVSVATSNGTESGAGSGTGTSTLTGNGTSGGTAGTTSSGTGAGG 420  
 QY 421 TSGTTTSSGAASGAGTGAGTTTSSGAGSDKAGTGTSTLTGNGTSGGTAGTTSSGTGAGG 480  
 DB 421 TSGTTTSSGAASGAGTGAGTTTSSGAGSDKAGTGTSTLTGNGTSGGTAGTTSSGTGAGG 480  
 QY 481 ASNAKIPGIMTLTLFALLTFIVN 503  
 DB 481 ASNAKIPGIMTLTLFALLTFIVN 503

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:36:24 ; Search time 140.806 Seconds  
(without alignments)  
476.010 Million cell updates/sec

Title: US-09-853-079-52  
Perfect score: 2565  
Sequence: 1 KRFPHTDMNGIHYIIDGS.....AKIPGIMTLTIFALLTFIVN 503

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2565	100.0	503	19	Babesia microti MN
2	2565	100.0	503	20	Babesia microti an
3	2565	100.0	503	21	B. microti clone a
4	2565	100.0	503	23	Babesia microti an
5	2239.5	87.3	452	23	BAB89024
6	2130	83.0	492	19	AAW56281
7	2130	83.0	492	20	AAW56281
8	2130	83.0	492	21	AAW56281
9	2130	83.0	492	21	AAW56281
10	1159.5	45.2	1132	21	AAW56281

11	1159.5	45.2	1132	23	ABB88976
12	1143	44.6	666	21	AAW56302
13	1143	44.6	666	23	ABB88975
14	1141	44.5	677	23	ABB88989
15	246.5	9.6	367	19	AAW56301
16	246.5	9.6	367	20	AAW56301
17	246.5	9.6	367	21	AAW56301
18	246.5	9.6	367	23	ABB88960
19	239.5	9.3	367	21	ABB88976
20	239.5	9.3	367	22	ABB88976
21	239.5	9.3	367	20	AAW56279
22	239.5	9.3	367	21	AAW56279
23	239.5	9.3	367	21	AAW56279
24	227.5	8.9	303	23	ABB88936
25	227.5	8.9	303	23	ABB88994
26	222.5	8.7	309	23	ABB88978
27	219	8.5	45	23	ABB88988
28	212	8.3	592	19	AAW56287
29	212	8.3	592	20	AAW56287
30	212	8.3	592	21	AAW56287
31	212	8.3	592	21	AAW56287
32	206.5	8.1	54	23	ABB88940
33	196.5	7.7	1175	21	AAW56287
34	196.5	7.7	1218	21	AAW56287
35	196.5	7.7	1218	22	ABB59285
36	194.5	7.6	1038	17	AAW56287
37	194.5	7.6	1038	22	AAW56287
38	194.5	7.6	1038	22	AAW56287
39	192	7.5	1011	21	AAW56287
40	192	7.5	1170	18	AAW56287
41	192	7.5	1170	19	AAW56287
42	191	7.4	231	23	AAW56287
43	191	7.4	463	19	AAW56287
44	191	7.4	463	20	AAW56287
45	191	7.4	463	21	AAW56287

ALIGNMENTS

RESULT 1  
AAW56302  
ID AAW56302 standard; Protein; 503 AA.  
XX  
AC AAW56302;  
XX  
DT 28-SEP-1998 (first entry)  
XX  
DE Babesia microti MN-10 antigen sequence.  
XX  
KW antigen; detection; diagnosis; vaccine; tick-borne disease;  
KW differentiation; Lyme disease; ehrlichiosis.  
XX  
OS Babesia microti.  
XX  
PN EP834567-A2.  
XX  
PD 08-APR-1998.  
XX  
PF 01-OCT-1997; 97EP-0117067.  
XX  
PR 24-APR-1997; 97US-0845258.  
PR 01-OCT-1996; 96US-0723142.  
XX  
XX (CORI-) CORIXA CORP.  
XX Houghton R, Lodes MJ, Reed SG, Sleath PR;  
DR WPI: 1998-195465/18.  
DR N-PSDB; AAW22748.  
XX  
PT Polypeptides comprising Babesia microti antigens and their immunogenic fragments or epitopes - and related nucleic acid,



```
XX 15-NOV-2001.
XX 09-MAY-2001; 2001WO-US15192.
XX 10-MAY-2000; 2000US-0569098.
XX 27-JUN-2000; 2000US-0605724.
XX 07-SEP-2000; 2000US-0656688.
XX 10-OCT-2000; 2000US-0685436.
XX 13-DEC-2000; 2000US-0737178.
XX 26-FEB-2001; 2001US-0794764.
XX (CORI-) CORIXA CORP.
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;
XX Secrist H;
XX WPI; 2002-216691/27.
XX New Babesia microti antigens, useful for diagnosing and treating B.
XX microti infection, and as component of a composition for enhancing
XX immune response against B. microti infections
XX Claim 35; Page 160-163; 195pp; English.
XX The present invention relates to novel Babesia microti antigens and their
XX coding sequences. The B. microti antigens, antigenic epitopes of such
XX antigens, and compositions comprising such antigens are useful for
XX diagnosing and treating B. microti infection. The compositions are
XX especially useful for enhancing immune response against B. microti
XX infection. The present sequence was used to illustrate the invention.
XX Sequence 677 AA;
XX Query Match 100.0%; Score 139; DB 23; Length 677;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-12;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX 1 GKPNTNKSEKAERKSHDTQTTOEICE 26
XX 175 GKPNTNKSEKAERKSHDTQTTOEICE 200
XX
XX RESULT 15
XX AAB30231
XX ID AAB30231 standard; Protein; 1132 AA.
XX AC AAB30231;
XX 12-FEB-2001 (first entry)
XX B. microti MN-10/BMNI-17/BMNI-15 fusion protein SEQ ID NO: 87.
XX Babesiosis; rodent parasite; tick-borne illness; antigen;
XX disease diagnosis; disease prevention.
XX Babesia sp.
XX Synthetic.
XX WO200060090-A1.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-US09136.
XX 05-APR-1999; 99US-0286488.
XX 17-MAR-2000; 2000US-0528784.
XX (CORI-) CORIXA CORP.
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;
XX WPI; 2000-686939/67.
XX
XX New polypeptides containing an antigenic portion of Babesia microti
XX antigen and DNAs encoding the polypeptides, useful for diagnosing,
XX treating or preventing B. microti infection, or for inducing protective
XX immunity in a patient
XX Example 7; Page 112-116; 118pp; English.
XX The present invention is related to the isolation of antigenic sequences
XX from the rodent parasite Babesia microti. This organism is transmitted to
XX humans by the same tick which transmits Lyme disease and ehrlichiosis.
XX The organism causes a malaria-like infection known as babesiosis. The
XX sequences identified by this invention can be used in the diagnosis,
XX prevention and treatment of babesiosis.
XX Sequence 1132 AA;
XX Query Match 100.0%; Score 139; DB 21; Length 1132;
XX Best Local Similarity 100.0%; Pred. No. 2.1e-12;
XX Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX 1 GKPNTNKSEKAERKSHDTQTTOEICE 26
XX 869 GKPNTNKSEKAERKSHDTQTTOEICE 894
XX
XX Search completed: July 16, 2003, 17:46:46
XX Job time : 8.27825 secs
```

CC coding sequences. The B. microti antigens, antigenic epitopes of such  
 CC antigens, and compositions comprising such antigens are useful for  
 CC diagnosing and treating B. microti infection. The compositions are  
 CC especially useful for enhancing immune response against B. microti  
 CC infection. The present sequence was used to illustrate the invention.

XX Sequence 481 AA;

Query Match 100.0%; Score 139; DB 23; Length 481;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-13;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPNINKSEAKERSHDTQTTOEICE 26  
 Db 218 GKPNINKSEAKERSHDTQTTOEICE 243

# RESULT 12

AAB30230  
 ID AAB30230 standard; Protein; 666 AA.

XX AAB30230;

DT 12-FEB-2001 (first entry)

DE B. microti MN-10/BMNI-17 fusion protein SEQ ID NO: 85.

KW Babesiosis; rodent parasite; tick-borne illness; antigen;  
 KW disease diagnosis; disease prevention.

OS Babesia sp.

OS Synthetic.

PN WO200060090-A1.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-US09136.

PR 05-APR-1999; 99US-0286488.

PR 17-MAR-2000; 2000US-0528784.

PA (CORI-) CORIXA CORP.

PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;

DR WPI; 2000-686939/67.

XX New polypeptides containing an antigenic portion of Babesia microti  
 PT antigen and DNAs encoding the polypeptides, useful for diagnosing,  
 PT treating or preventing B. microti infection, or for inducing protective  
 PT immunity in a patient.

PS Example 7; Page 108-111; 118pp; English.

CC The present invention is related to the isolation of antigenic sequences  
 CC from the rodent parasite Babesia microti. This organism is transmitted to  
 CC humans by the same tick which transmits Lyme disease and ehrlichiosis.  
 CC The organism causes a malaria-like infection known as babesiosis. The  
 CC sequences identified by this invention can be used in the diagnosis,  
 CC prevention and treatment of babesiosis.

XX Sequence 666 AA;

Query Match 100.0%; Score 139; DB 21; Length 666;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-12;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPNINKSEAKERSHDTQTTOEICE 26  
 Db 403 GKPNINKSEAKERSHDTQTTOEICE 428

# RESULT 13

ABB88975

ID ABB88975 standard; Protein; 666 AA.

XX ABB88975;

DT 20-JUN-2002 (first entry)

DE Babesia microti antigenic epitope fusion protein Baf-3.

KW Protozoacide; vaccine; antigen; antigenic epitope; infection.

OS Babesia microti.

PN WO200185947-A2.

PD 15-NOV-2001.

PF 09-MAY-2001; 2001WO-US15192.

PR 10-MAY-2000; 2000US-0569098.

PR 27-JUN-2000; 2000US-0605724.

PR 07-SEP-2000; 2000US-0656688.

PR 10-OCT-2000; 2000US-0685436.

PR 13-DEC-2000; 2000US-0737178.

PR 26-FEB-2001; 2001US-0794764.

PA (CORI-) CORIXA CORP.

PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;

PI Secrist H;

DR WPI; 2002-216691/27.

XX New Babesia microti antigens, useful for diagnosing and treating B.

PT microti infection, and as component of a composition for enhancing

PT immune response against B. microti infections

PS Claim 35; Page 113-115; 195pp; English.

XX The present invention relates to novel Babesia microti antigens and their

CC coding sequences. The B. microti antigens, antigenic epitopes of such

CC antigens, and compositions comprising such antigens are useful for

CC diagnosing and treating B. microti infection. The compositions are

CC especially useful for enhancing immune response against B. microti

CC infection. The present sequence was used to illustrate the invention.

XX Sequence 666 AA;

Query Match 100.0%; Score 139; DB 23; Length 666;

Best Local Similarity 100.0%; Pred. No. 1.1e-12;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKPNINKSEAKERSHDTQTTOEICE 26

Db 403 GKPNINKSEAKERSHDTQTTOEICE 428

# RESULT 14

ABB88989

ID ABB88989 standard; Protein; 677 AA.

XX ABB88989;

DT 20-JUN-2002. (first entry)

DE Babesia microti antigenic epitope fusion protein Baf-5.

KW Protozoacide; vaccine; antigen; antigenic epitope; infection.

OS Babesia microti.

PN WO200185947-A2.

DE B. microti BMN1-17 antigen reverse complement SEQ ID NO: 38.  
 XX Babesiosis; rodent parasite; tick-borne illness; antigen;  
 KW disease diagnosis; disease prevention.  
 XX

OS Babesia microti.  
 XX

PN WO200060090-A1.  
 XX

PD 12-OCT-2000.  
 XX

XX 05-APR-2000; 2000WO-US09136.  
 PF

XX 05-APR-1999; 99US-0286488.  
 PR

PR 17-MAR-2000; 2000US-0528784.  
 XX

PA (CORI-) CORIXA CORP.  
 XX

XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD;  
 PI WPI; 2000-686939/67.  
 XX

XX New polypeptides containing an antigenic portion of Babesia microti  
 XX antigen and DNAs encoding the polypeptides, useful for diagnosing,  
 PT treating or preventing B. microti infection, or for inducing protective  
 PT immunity in a patient -  
 XX

XX Example 1; Page 86-87; 118pp; English.  
 PS

XX The present invention is related to the isolation of antigenic sequences  
 CC from the rodent parasite Babesia microti. This organism is transmitted to  
 CC humans by the same tick which transmits Lyme disease and ehrlichiosis.  
 CC The organism causes a malaria-like infection known as babesiosis. The  
 CC sequences identified by this invention can be used in the diagnosis,  
 CC prevention and treatment of babesiosis.  
 XX

SQ Sequence 445 AA;  
 XX

Query Match 100.0%; Score 139; DB 21; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-13;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GKPNTNKSEKAEKSHDTQTTOEICE 26  
 DB 182 GKPNTNKSEKAEKSHDTQTTOEICE 207  
 |||||||||||||||||||

RESULT 10  
 ABB88952  
 ID ABB88952 standard; Protein; 445 AA.  
 XX

AC ABB88952;  
 XX

XX 20-JUN-2002 (first entry)  
 DT

XX Babesia microti antigen epitope #1.  
 DE

XX Protozoacide; vaccine; antigen; antigenic epitope; infection.  
 KW

XX Babesia microti.  
 OS

XX WO200185947-A2.  
 PN

XX 15-NOV-2001.  
 PD

XX 09-MAY-2001; 2001WO-US15192.  
 PF

XX 10-MAY-2000; 2000US-0569098.  
 PR

PR 27-JUN-2000; 2000US-0605724.  
 PR

PR 07-SEP-2000; 2000US-0656688.  
 PR

PR 10-OCT-2000; 2000US-0685436.  
 PR

PR 13-DEC-2000; 2000US-0737178.  
 PR

PR 26-FEB-2001; 2001US-0794764.  
 PR

XX (CORI-) CORIXA CORP.  
 PA

XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;  
 PI Secrist H;  
 PI

XX WPI; 2002-216691/27.  
 DR

XX New Babesia microti antigens, useful for diagnosing and treating B.  
 XX microti infection, and as component of a composition for enhancing  
 PT immune response against B. microti infections -  
 PT

XX Example 1; Page 92-93; 195pp; English.  
 PS

XX The present invention relates to novel Babesia microti antigens and their  
 CC coding sequences. The B. microti antigens, antigenic epitopes of such  
 CC antigens, and compositions comprising such antigens are useful for  
 CC diagnosing and treating B. microti infection. The compositions are  
 CC especially useful for enhancing immune response against B. microti  
 CC infection. The present sequence was used to illustrate the invention.  
 XX

SQ Sequence 445 AA;  
 XX

Query Match 100.0%; Score 139; DB 23; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-13;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GKPNTNKSEKAEKSHDTQTTOEICE 26  
 DB 182 GKPNTNKSEKAEKSHDTQTTOEICE 207  
 |||||||||||||||||||

RESULT 11  
 ABB89025  
 ID ABB89025 standard; Protein; 481 AA.  
 XX

AC ABB89025;  
 XX

XX 20-JUN-2002 (first entry)  
 DT

XX Babesia microti antigenic epitope fusion protein.  
 DE

XX Protozoacide; vaccine; antigen; antigenic epitope; infection.  
 KW

XX Babesia microti.  
 OS

XX WO200185947-A2.  
 PN

XX 15-NOV-2001.  
 PD

XX 09-MAY-2001; 2001WO-US15192.  
 PF

XX 10-MAY-2000; 2000US-0569098.  
 PR

PR 27-JUN-2000; 2000US-0605724.  
 PR

PR 07-SEP-2000; 2000US-0656688.  
 PR

PR 10-OCT-2000; 2000US-0685436.  
 PR

PR 13-DEC-2000; 2000US-0737178.  
 PR

PR 26-FEB-2001; 2001US-0794764.  
 PR

XX (CORI-) CORIXA CORP.  
 PA

XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;  
 PI Secrist H;  
 PI

XX WPI; 2002-216691/27.  
 DR

XX New Babesia microti antigens, useful for diagnosing and treating B.  
 XX microti infection, and as component of a composition for enhancing  
 PT immune response against B. microti infections -  
 PT

XX Claim 2; Page 191-192; 195pp; English.  
 PS

XX The present invention relates to novel Babesia microti antigens and their



XX PS Example 1; Page 103; 195pp; English.  
 CC The present invention relates to novel Babesia microti antigens and their  
 CC coding sequences. The B. microti antigens, antigenic epitopes of such  
 CC antigens, and compositions comprising such antigens are useful for  
 CC diagnosing and treating B. microti infection. The compositions are  
 CC especially useful for enhancing immune response against B. microti  
 CC infection. The present sequence was used to illustrate the invention.  
 XX SQ Sequence 275 AA;

Query Match 100.0%; Score 139; DB 23; Length 275;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-13;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GKPTNKSEKAERKSHDTQTTOEICE 26  
 Db 218 GKPTNKSEKAERKSHDTQTTOEICE 243

RESULT 7  
 AAW56298  
 ID AAW56298 standard; Protein: 445 AA.  
 AC AAW56298;  
 XX 28-SEP-1998 (first entry)  
 DT Babesia microti BMNI-17 complement antigen sequence.  
 DE antigen; detection; diagnosis; vaccine; tick-borne disease;  
 KW differentiation; Lyme disease; ehrlichiosis.  
 KW Babesia microti.  
 OS EP834567-A2.  
 XX 08-APR-1998.  
 PD 01-OCT-1997; 97EP-0117067.  
 PF 24-APR-1997; 97US-0845258.  
 PR 01-OCT-1996; 96US-0723142.  
 XX (CORI-) CORIXA CORP.  
 PA Houghton R, Lodes MJ, Reed SG, Sleath PR;  
 PI WPI; 1998-195465/18.  
 DR N-PSDB; AAW22753.  
 XX Polypeptides comprising Babesia microti antigens and their  
 PT immunogenic fragments or epitopes - and related nucleic acid,  
 PT vectors, transformed cells and antibodies, useful for diagnosis of  
 PT infection and in protective vaccines

XX PS Claim 1; Page 77-79; 113pp; English.  
 CC The sequence is that of a polypeptide comprising at least  
 CC one antigenic portion of a Babesia microti antigen. It can be used  
 CC to diagnose B. microti infection by detecting specific antibodies  
 CC in usual immunoassays. Infection can also be diagnosed using:  
 CC (a) primers or probes derived from the coding sequence, in  
 CC standard amplification or hybridisation tests, or (b) using  
 CC antibodies to detect the corresponding antigen. It is also  
 CC useful in vaccines to protect against infection, especially  
 CC when formulated with an adjuvant. The new diagnostic methods  
 CC allow rapid differentiation between B. microti infection and  
 CC other tick-borne diseases (Lyme disease and ehrlichiosis) that  
 CC have similar symptoms but require different treatments.

XX SQ Sequence 445 AA;

Query Match 100.0%; Score 139; DB 19; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-13;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GKPTNKSEKAERKSHDTQTTOEICE 26  
 Db 182 GKPTNKSEKAERKSHDTQTTOEICE 207

RESULT 8  
 AAY24358  
 ID AAY24358 standard; Protein: 445 AA.  
 AC AAY24358;  
 XX 16-SEP-1999 (first entry)  
 DT Babesia microti antigen BMNI-17 complementary open reading frame protein.  
 DE Babesia microti; antigen; immunogen; diagnosis; infection; vaccine;  
 KW immunity; detection.  
 KW Babesia microti.  
 OS WO9929869-A1.  
 XX 17-JUN-1999.  
 PD 11-DEC-1998; 98WO-US26437.  
 PF 11-DEC-1997; 97US-0990571.  
 PR (CORI-) CORIXA CORP.  
 PA (MAYO-) MAYO FOUNDATION.  
 XX Bruinsma E, Houghton R, Lodes MJ, Persing D, Reed SG;  
 PI Sleath PR;  
 XX WPI; 1999-385612/32.  
 DR N-PSDB; AAX90012.  
 XX New isolated Babesia microti polypeptides  
 PT Example 1; Page 91-92; 126pp; English.

XX CC The present invention describes isolated polypeptides comprising  
 CC specific immunogenic portions of Babesia microti. AAX88983 to AAX88994  
 CC encode specifically claimed B. microti immunogenic proteins, and  
 CC AAY24327 to AAY24338 represent these proteins. B. microti polypeptides  
 CC and nucleic acids can be used for detecting B. microti infections. They  
 CC can also be used in vaccines for inducing protective immunity against B.  
 CC microti infections. The present sequence encodes a B. microti antigen.

XX SQ Sequence 445 AA;

Query Match 100.0%; Score 139; DB 20; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-13;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GKPTNKSEKAERKSHDTQTTOEICE 26  
 Db 182 GKPTNKSEKAERKSHDTQTTOEICE 207

RESULT 9  
 AAB30207  
 ID AAB30207 standard; Protein: 445 AA.  
 AC AAB30207;  
 XX 12-FEB-2001 (first entry)  
 DT

```

AA24365
ID  AAY24365 standard; Protein; 275 AA.
XX
AC  AAY24365;
XX
DT  16-SEP-1999 (first entry)
XX
DE  Babesia microti antigen BMNI-20 complementary open reading frame protein.
XX
KW  Babesia microti; antigen; immunogen; diagnosis; infection; vaccine;
KW  immunity; detection.
XX
OS  Babesia microti.
XX
PN  WO9929869-A1.
XX
PD  17-JUN-1999.
XX
PF  11-DEC-1998; 98WO-US26437.
XX
PR  11-DEC-1997; 97US-0990571.
XX
PA  (CORI-) CORIXA CORP.
PA  (MAYO-) MAYO FOUNDATION.
XX
PI  Bruinsma E, Houghton R, Lodes MJ, Persing D, Reed SG;
PI  Sleath PR;
XX
DR  WPI; 1999-385612/32.
DR  N-PSDB; AAX90017.
XX
XX  New isolated Babesia microti polypeptides
XX
PS  Example 1; Page 109-110; 126pp; English.
XX
CC  The present invention describes isolated polypeptides comprising
CC  specific immunogenic portions of Babesia microti. AAX88983 to AAX88994
CC  encode specifically claimed B. microti immunogenic proteins, and
CC  AAY24327 to AAY24338 represent these proteins. B. microti polypeptides
CC  and nucleic acids can be used for detecting B. microti infections. They
CC  can also be used in vaccines for inducing protective immunity against B.
CC  microti infections. The present sequence represents a B. microti antigen.
XX
SQ  Sequence 275 AA;
XX
Query Match 100.0%; Score 139; DB 20; Length 275;
Best Local Similarity 100.0%; Pred. No. 4.1e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 GKPNTNKSEKAERKSHDTQTQTEICE 26
Db  218 GKPNTNKSEKAERKSHDTQTQTEICE 243

RESULT 5
AAB30217
ID  AAB30217 standard; Protein; 275 AA.
XX
AC  AAB30217;
XX
DT  12-FEB-2001 (first entry)
XX
DE  B. microti clone antigen SEQ ID NO: 53.
XX
W  Babesiosis; rodent parasite; tick-borne illness; antigen;
W  disease diagnosis; disease prevention.
XX
XX  'a microti.
XX  -A1.
XX  .0.

Query Match 100.0%; Score 139; DB 20; Length 275;
Best Local Similarity 100.0%; Pred. No. 4.1e-13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1 GKPNTNKSEKAERKSHDTQTQTEICE 26
Db  218 GKPNTNKSEKAERKSHDTQTQTEICE 243

RESULT 6
ABB88962
ID  ABB88962 standard; Protein; 275 AA.
XX
AC  ABB88962;
XX
DT  20-JUN-2002 (first entry)
XX
DE  Babesia microti antigen BMNI-20.
XX
KW  Protozoacide; vaccine; antigen; antigenic epitope; infection.
XX
OS  Babesia microti.
XX
PN  WO200185947-A2.
XX
PD  15-NOV-2001.
XX
PF  09-MAY-2001; 2001WO-US15192.
XX
PR  10-MAY-2000; 2000US-0569098.
PR  27-JUN-2000; 2000US-0605724.
PR  07-SEP-2000; 2000US-0656688.
PR  10-OCT-2000; 2000US-0685436.
PR  13-DEC-2000; 2000US-0737178.
PR  26-FEB-2001; 2001US-0794764.
XX
PA  (CORI-) CORIXA CORP.
XX
PI  Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;
PI  Secrist H;
XX
DR  WPI; 2002-216691/27.
XX
PT  New Babesia microti antigens, useful for diagnosing and treating B.
PT  microti infection, and as component of a composition for enhancing
PT  immune response against B. microti infections

```

XX New Babesia microti antigens, useful for diagnosing and treating B.  
PT microti infection, and as component of a composition for enhancing  
PT immune response against B. microti infections  
XX  
PS Claim 2; Page 185; 195pp; English.  
XX  
CC The present invention relates to novel Babesia microti antigens and their  
CC coding sequences. The B. microti antigens, antigenic epitopes of such  
CC antigens, and compositions comprising such antigens are useful for  
CC diagnosing and treating B. microti infection. The compositions are  
CC especially useful for enhancing immune response against B. microti  
CC infection. The present sequence was used to illustrate the invention.  
XX  
SQ Sequence 26 AA;  
Query Match 100.0%; Score 139; DB 23; Length 26;  
Best Local Similarity 100.0%; Pred. No. 2.8e-14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GKPNTNKSEKAERKSHDTQTQICE 26  
DB 1 GKPNTNKSEKAERKSHDTQTQICE 26  
|||||  
RESULT 2  
ABB89012  
ID ABB89012 standard; Protein; 50 AA.  
XX  
AC ABB89012;  
XX  
XX 20-JUN-2002 (first entry)  
DE Babesia microti antigen peptide.  
XX  
XX Protozoacide; vaccine; antigen; antigenic epitope; infection.  
XX  
OS Babesia microti.  
XX  
XX WO200185947-A2.  
XX  
XX 15-NOV-2001.  
XX  
XX 09-MAY-2001; 2001WO-US15192.  
XX  
XX 10-MAY-2000; 2000US-0569098.  
XX  
XX 27-JUN-2000; 2000US-0605724.  
XX  
XX 07-SEP-2000; 2000US-0656688.  
XX  
XX 10-OCT-2000; 2000US-0685436.  
XX  
XX 13-DEC-2000; 2000US-0737178.  
XX  
XX 26-FEB-2001; 2001US-0794764.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;  
PI Secret H;  
XX  
XX WPI; 2002-216691/27.  
XX  
XX New Babesia microti antigens, useful for diagnosing and treating B.  
PT microti infection, and as component of a composition for enhancing  
PT immune response against B. microti infections  
XX  
XX Claim 2; Page 185; 195pp; English.  
XX  
CC The present invention relates to novel Babesia microti antigens and their  
CC coding sequences. The B. microti antigens, antigenic epitopes of such  
CC antigens, and compositions comprising such antigens are useful for  
CC diagnosing and treating B. microti infection. The compositions are  
CC especially useful for enhancing immune response against B. microti  
CC infection. The present sequence was used to illustrate the invention.  
XX  
SQ Sequence 50 AA;

Query Match 100.0%; Score 139; DB 23; Length 50;  
Best Local Similarity 100.0%; Pred. No. 5.9e-14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GKPNTNKSEKAERKSHDTQTQICE 26  
DB 1 GKPNTNKSEKAERKSHDTQTQICE 26  
|||||  
RESULT 3  
AAW56303  
ID AAW56303 standard; Protein; 275 AA.  
XX  
AC AAW56303;  
XX  
XX 28-SEP-1998 (first entry)  
XX  
DE Babesia microti BMNI-20 antigen sequence.  
XX  
XX antigen; detection; diagnosis; vaccine; tick-borne disease;  
KW differentiation; Lyme disease; ehrlichiosis.  
XX  
XX Babesia microti.  
XX  
XX EP834567-A2.  
XX  
XX 08-APR-1998.  
XX  
XX 01-OCT-1997; 97EP-0117067.  
XX  
XX 24-APR-1997; 97US-0845258.  
XX  
XX 01-OCT-1996; 96US-0723142.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Houghton R, Lodes MJ, Reed SG, Sleath PR;  
PI WPI; 1998-195465/18.  
XX  
XX N-PSDB; AAV22749.  
XX  
XX Polypeptides comprising Babesia microti antigens and their  
PT immunogenic fragments or epitopes - and related nucleic acid,  
PT vectors, transformed cells and antibodies, useful for diagnosis of  
PT infection and in protective vaccines  
XX  
XX Claim 1; Page 101-102; 113pp; English.  
XX  
CC The sequence is that of a polypeptide comprising at least  
CC one antigenic portion of a Babesia microti antigen. It can be used  
CC to diagnose B. microti infection by detecting specific antibodies  
CC in usual immunoassays. Infection can also be diagnosed using:  
CC (a) primers or probes derived from the coding sequence, in  
CC standard amplification or hybridisation tests, or (b) using  
CC antibodies to detect the corresponding antigen. It is also  
CC useful in vaccines to protect against infection, especially  
CC when formulated with an adjuvant. The new diagnostic methods  
CC allow rapid differentiation between B. microti infection and  
CC other tick-borne diseases (Lyme disease and ehrlichiosis) that  
CC have similar symptoms but require different treatments.  
XX  
SQ Sequence 275 AA;  
Query Match 100.0%; Score 139; DB 19; Length 275;  
Best Local Similarity 100.0%; Pred. No. 4.1e-13;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GKPNTNKSEKAERKSHDTQTQICE 26  
DB 218 GKPNTNKSEKAERKSHDTQTQICE 243  
|||||

RESULT 4

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 16, 2003, 17:36:24 ; Search time 7.27825 Seconds  
(without alignments)  
476.010 Million cell updates/sec

Title: US-09-853-079-195

Perfect score: 139

Sequence: 1 GRPNTNKSEAKRSHDTQTQICE 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	139	100.0	26	ABB89013	Babesia microti an
2	139	100.0	50	ABB89012	Babesia microti an
3	139	100.0	275	AAW56303	Babesia microti BM
4	139	100.0	275	AAAY24365	Babesia microti an
5	139	100.0	275	AAAY30217	B. microti clone a
6	139	100.0	275	ABB89962	Babesia microti an
7	139	100.0	445	AAW56298	Babesia microti BM
8	139	100.0	445	AAAY24358	Babesia microti an
9	139	100.0	445	AAAY30207	B. microti BMNI-17
10	139	100.0	445	ABB88952	Babesia microti an

11	139	100.0	481	23	ABB89025	Babesia microti an
12	139	100.0	666	21	AAAB30230	B. microti MN-10/B
13	139	100.0	666	23	ABB88975	Babesia microti an
14	139	100.0	677	23	ABB88989	Babesia microti an
15	139	100.0	1132	21	AAAB30231	B. microti MN-10/B
16	139	100.0	1132	23	ABB88976	Babesia microti an
17	115	82.7	25	23	ABB89019	Babesia microti an
18	78	56.1	25	23	ABB89015	Babesia microti an
19	65	46.8	25	23	ABB89010	Babesia microti an
20	63	45.3	32	20	AAAY24359	Babesia microti an
21	63	45.3	32	21	AAAB30208	B. microti BMNI-17
22	63	45.3	32	23	ABB88953	Babesia microti an
23	54	38.8	600	22	AAAG75092	Human colon cancer
24	54	38.8	1048	21	AAAY68823	Amino acid sequenc
25	54	38.8	1158	23	ABB97465	Novel human protei
26	51	36.7	25	23	ABB89020	Babesia microti an
27	51	36.7	212	20	AAAY35210	Chlamydia pneumoni
28	51	36.7	651	23	ABB90561	Chlamydia pneumoni
29	51	36.7	651	23	AAE20304	Chlamydia pneumoni
30	51	36.7	672	23	AAE20288	Chlamydia pneumoni
31	50	36.0	584	20	AAAY34934	Chlamydia pneumoni
32	49.5	35.6	523	21	AAAY56856	Morbilivirus anti
33	49.5	35.6	776	22	ABG29124	Novel human diagno
34	49	35.3	25	23	ABB89018	Babesia microti an
35	49	35.3	278	21	AAAB56745	Human prostate can
36	49	35.3	400	22	ABB68410	Drosophila melanog
37	48	34.5	149	21	AAAG27807	Arabidopsis thalia
38	48	34.5	188	21	AAAG27806	Arabidopsis thalia
39	47.5	34.2	474	22	ABB62780	Drosophila melanog
40	47.5	34.2	474	22	ABB66823	Drosophila melanog
41	47	33.8	83	21	AAAB34321	Human secreted pro
42	47	33.8	99	22	AAAB39757	Human polypeptide
43	47	33.8	131	23	ABP42747	Human ovarian anti
44	47	33.8	154	21	AAAB25692	Human secreted pro
45	47	33.8	178	20	AAAY41674	Human channel-rela

#### ALIGNMENTS

RESULT 1  
ABB89013  
ID ABB89013 standard; Peptide; 26 AA.  
XX  
AC ABB89013;  
XX  
DT 20-JUN-2002 (first entry)  
XX  
DE DE  
XX  
XX  
KW Protozoacide; vaccine; antigen; antigenic epitope; infection.  
XX  
OS Babesia microti.  
XX  
XX  
PN WO200185947-A2.  
XX  
PD 15-NOV-2001.  
XX  
PF 09-MAY-2001; 2001WO-US15192.  
XX  
PR 10-MAY-2000; 2000US-0569098.  
PR 27-JUN-2000; 2000US-0605724.  
PR 07-SEP-2000; 2000US-0656688.  
PR 10-OCT-2000; 2000US-0685436.  
PR 13-DEC-2000; 2000US-0737178.  
PR 26-FEB-2001; 2001US-0794764.  
(CORI-) CORIXA CORP.  
PI Reed SG, Lodes MJ, Houghton RL, Sleath PR, McNeill PD, Homer MJ;  
PI Secrist H;  
XX  
DR WPI; 2002-216691/27.